Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **502,** 4328–4349 (2021) doi:10.1093/mnras/stab292
Advance Access publication 2021 February 3

Characterizing the target selection pipeline for the Dark Energy Spectroscopic Instrument Bright Galaxy Survey

Omar Ruiz-Macias [1,2]★ Pauline Zarrouk,[1] Shaun Cole,[1] Carlton M. Baugh,[1,2] Peder Norberg,[1,3] John Lucey [3] Arjun Dey,[4] Daniel J. Eisenstein,[5] Peter Doel,[6] Enrique Gaztañaga [7] ChangHoon Hahn [8,9] Robert Kehoe,[10] Ellie Kitanidis,[11] Martin Landriau,[8] Dustin Lang,[12,13] John Moustakas,[14] Adam D. Myers,[15] Francisco Prada,[16] Michael Schubnell,[17] David H. Weinberg[18] and M. J. Wilson[8,9]

[1]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[2]*Institute for Data Science, Durham University, South Road, Durham DH1 3LE, UK*
[3]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[4]*NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave., Tucson, AZ 85719, USA*
[5]*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[6]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[7]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Bellaterra (Barcelona), Spain*
[8]*Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA*
[9]*Berkeley Center for Cosmological Physics, UC Berkeley, Berkeley, CA 94720, USA*
[10]*Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75205, USA*
[11]*Department of Physics, University of California-Berkeley, 366 LeConte Hall, Berkeley, CA 94720, USA*
[12]*Perimeter Institute for Theoretical Physics, 31 Caroline Street N, Waterloo, ON N2L 2Y5, Canada*
[13]*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[14]*Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA*
[15]*University of Wyoming, 1000 E. University Ave., Laramie, WY 82071, USA*
[16]*Instituto de Astrofisica de Andalucía, Glorieta de la Astronomía, s/n, E-18008 Granada, Spain*
[17]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[18]*Department of Astronomy and the Center for Cosmology and Astroparticle Physics, The Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA*

Accepted 2021 January 26. Received 2021 January 20; in original form 2020 July 30

ABSTRACT

We present the steps taken to produce a reliable and complete input galaxy catalogue for the Dark Energy Spectroscopic Instrument (DESI) Bright Galaxy Survey (BGS) using the photometric Legacy Survey DR8 DECam. We analyse some of the main issues faced in the selection of targets for the DESI BGS, such as star–galaxy separation, contamination by fragmented stars and bright galaxies. Our pipeline utilizes a new way to select BGS galaxies using *Gaia* photometry and we implement geometrical and photometric masks that reduce the number of spurious objects. The resulting catalogue is cross-matched with the Galaxy And Mass Assembly (GAMA) survey to assess the completeness of the galaxy catalogue and the performance of the target selection. We also validate the clustering of the sources in our BGS catalogue by comparing with mock catalogues and the Sloan Digital Sky Survey (SDSS) data. Finally, the robustness of the BGS selection criteria is assessed by quantifying the dependence of the target galaxy density on imaging and other properties. The largest systematic correlation we find is a 7 per cent suppression of the target density in regions of high stellar density.

Key words: catalogues – surveys – large-scale structure of Universe.

1 INTRODUCTION

The Dark Energy Spectroscopic Instrument[1] (DESI; DESI Collaboration et al. 2016) is a multifibre spectrograph that will be used to carry out a number of wide-field surveys of galaxies and quasars to map the large-scale structure of the Universe. These surveys will probe the form of dark energy by allowing high precision measurements of the baryon acoustic oscillation (BAO) scale and the growth rate of structure using redshift-space distortions (RSDs). The characterization and definition of the target list for each DESI survey is a critical step for efficient survey execution and to allow reliable measurements of galaxy clustering. Here we describe this

★ E-mail: omar.a.ruiz-macias@durham.ac.uk
[1]http://desi.lbl.gov/

process for the DESI Bright Galaxy Survey (BGS), a flux limited sample of around 10 million galaxies, using photometry from a new imaging survey, the Legacy Survey (LS).[2]

DESI is a robotically actuated, fibre-fed spectrograph that is capable of collecting 5000 spectra simultaneously.

The spectra cover the wavelength range 360–980 nm, with a spectral resolution of $R = \lambda/\Delta\lambda$ between 2000 and 5500, depending on the wavelength. DESI will be used to conduct a 5-yr survey starting in 2020, with the aim of measuring redshifts over a solid angle of 14 000 deg^2. More than 30 million spectroscopic targets will be selected for four different tracer samples drawn from the imaging data. These are (i) luminous red galaxies (LRGs) in the redshift range $z = 0.3$–1; (ii) emission line galaxies (ELGs) to $z = 1.7$; (iii) quasars to higher redshifts ($2.1 < z < 3.5$); and (iv) a magnitude-limited BGS out to $z \approx 0.6$ with a median redshift of $z \approx 0.2$ that is the focus of this paper.

DESI observations are divided into two main programmes: the Bright Time Survey (BTS) and the Dark Time Survey (DTS). The BGS will be part of the BTS and is conducted when the Moon is above the horizon and the sky is too bright to allow efficient observation of fainter targets. The BTS excludes the few nights closest to full Moon and BGS always targets fields that are at least 40°–50° away from the Moon. BGS alone will be 10 times larger than the Sloan Digital Sky Survey I (SDSS-I) and Sloan Digital Sky Survey II (SDSS-II) main galaxy samples (MGSs) of 1 million bright galaxies that were observed over the time period 1999–2008 (Abazajian et al. 2003).

The target sample for the BGS is intended to be a galaxy sample that is flux limited in the *r* band. The magnitude limit is determined by the total amount of bright observing time and the exposure times required to achieve the desired redshift efficiency. This target selection is, in essence, a deeper version of the target selection for the SDSS MGS (Strauss et al. 2002).

To make predictions for BGS target sample we make use of the mock galaxy catalogue created from the Millennium-XXL (MXXL) *N*-body simulation of Angulo et al. (2012) by Smith et al. (2017). This mock is tuned match the luminosity function, colour distribution, and clustering properties of the SDSS MGS at low redshift, and the evolution of these statistics to redshift $z \approx 0.5$ as measured from the Galaxy And Mass Assembly (GAMA) survey (Driver et al. 2011; Liske et al. 2015; Baldry et al. 2017).

The DESI BGS is expected to have a target density of just over 800 galaxies per deg^2 in a primary sample defined by a faint *r*-band magnitude limit of 19.5. Then, in a lower priority sample, a secondary sample of ∼600 galaxies deg^{-2} defined by the magnitude range $19.5 < r < 20$ (DESI Collaboration et al. 2016). From hereon we will refer to these BGS samples as BGS BRIGHT and BGS FAINT, respectively. A few per cent of galaxies in the DESI BGS will be lost due to deblending errors, superposition with bright stars, and other artefacts that typically affect imaging catalogues. Our aim is to provide a reliable input galaxy catalogue for the DESI BGS and to characterize its properties, such as the surface density of galaxies and their clustering. A complementary study by Kitanidis et al. (2020) examined the impact of imaging systematics on the selection and clustering of targets in the LRG, ELG, and quasi-stellar object (QSO) DESI surveys, using an earlier release of the LS imaging data (Dey et al. 2019).

Here, we define and characterized the BGS target selection based on the latest Dark Energy Camera Legacy Survey (DECaLS) release, DR8, which covers ∼2/3 of the full 14 000 deg^2 of DESI footprint. The resulting catalogue is defined in Ruiz-Macias et al. (2020) and here we present the details of that selection and associated analysis of the catalogue. This BGS catalogue was used by DESI in the commissioning stage of the early survey validation observations. It is planned that the final BGS catalogue will be based on the next, DR9, LS data release. This release will include better modelling of large galaxies and the light in bright star haloes. More discussion of DR9 and planned subsequent characterization of the BGS selection can be found in Section 6.

This paper is organized as follows. In Section 2, we describe the LSs imaging data used to select our targets and the secondary data sets used to tune the selection. In Sections 3 and 4, we define the spatial and photometric cuts used to select BGS targets and to get rid of artefacts that might become problematic for DESI observations plus the removal of poor quality imaging data. In Section 4, we define our star–galaxy classification using *Gaia* DR2. In Section 5, we compare the BGS catalogue with its overlap of the GAMA DR4[3] (Driver et al. 2011; Liske et al. 2015; Baldry et al. 2017) to assess the completeness and contamination of the BGS and to quantify its expected redshift distribution. In Section 5.2, we look at eight potential systematics that might be affecting our BGS target selection and try to mitigate these effects with linear weights determined using the stellar density. Section 5.3 shows the clustering of our BGS selection before and after applying the weights and we compare it with SDSS and the MXXL light-cone catalogue (Smith et al. 2017). Finally, in Section 6, we summarize our results and present our conclusions.

2 PHOTOMETRIC DATA SETS

During the BGS target selection process we make use of several catalogues. The main data set used is the Legacy Surveys DR8 (hereafter LSDR8) imaging catalogue from which we select our targets. We also make use of secondary catalogues for masking purposes, such as the Tycho-2 star catalogue (Høg et al. 2000), the *Gaia* DR2 (Gaia Collaboration et al. 2016a, 2018b), the Siena Galaxy Atlas – 2020 (SGA-2020; Moustakas, Lang, et al., in preparation), and globular clusters (GCs) from the OpenNGC[4] catalogue. We also use a combination of *Gaia* DR2 and LS photometry to perform star–galaxy separation.

2.1 Legacy Survey DR8 (DECam)

The Dark Energy Camera Legacy Survey (DECaLS), the Beijing–Arizona Sky Survey (BASS), and the Mayall *z*-band Legacy Survey (MzLS) together constitute the DESI Legacy Imaging Survey (hereafter the Legacy Survey). The imaging Legacy Survey was created with the aim of attaining photometry with the necessary target density, coverage, and depth required for DESI. The SDSS MGS (Strauss et al. 2002) and the Panoramic Survey Telescope and Rapid Response System 1 (Pan-STARRS1; Chambers et al. 2016)

[2]http://legacysurvey.org/

[3]This is an unreleased version of GAMA catalogue that the GAMA Collaboration made available to us. It is essentially the same as GAMA DR3, but with more redshifts.

[4]OpenNGC, https://github.com/mattiaverga/OpenNGC, is a data base containing positions and main data of New General Catalogue (NGC) and Index Catalogue (IC) objects constructed by the GAVO Data Center team by merging data from NASA/IPAC Extragalactic Database (NED), Hyper-Linked Extragalactic Databases and Archives (HyperLEDA), Set of Identifications, Measurements and Bibliography for Astronomical Data (SIMBAD), and several data bases available at High Energy Astrophysics Science Archive Research Center (HEASARC; https://heasarc.gsfc.nasa.gov/).

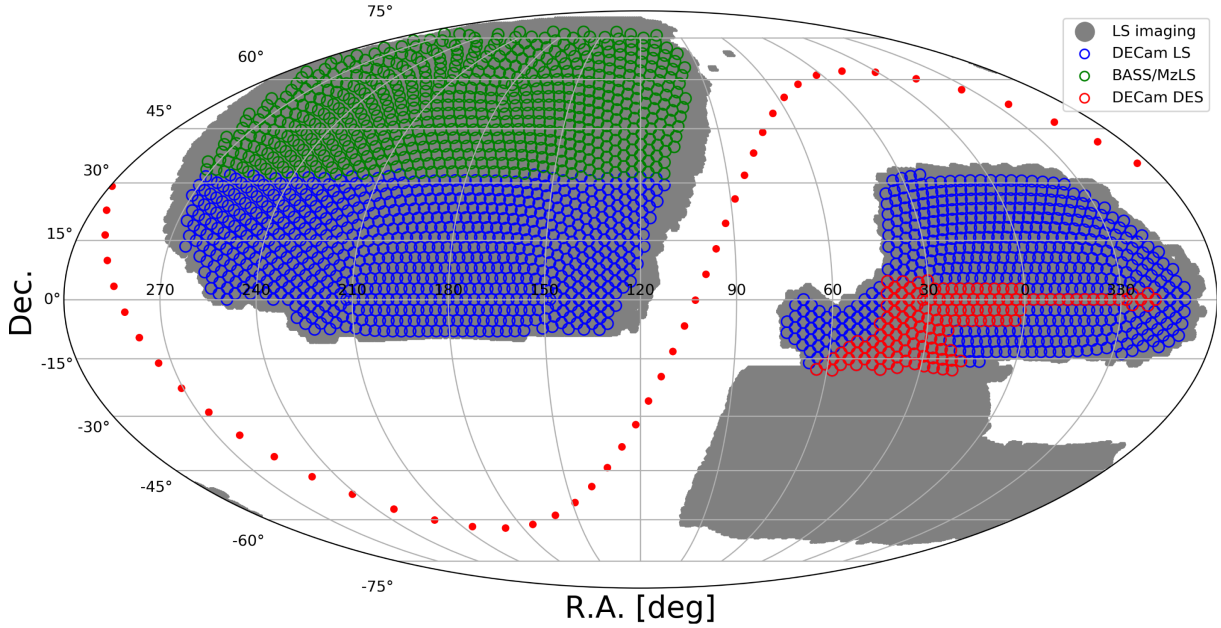

Figure 1. The sky map of the footprint of all the LS imaging used in DECaLS and in BASS and MzLS is shown in grey. The red and blue circles show the DESI tiles that define the portion of DESI survey footprint that lies within DECaLS. The blue tiles are those for which the data come from the DECam LS imaging, while the red tiles come from DECam DES imaging. The green tiles show the northern DESI footprint whose imaging data come from the BASS and MzLS surveys that are not the focus of this paper. The red dots show the locus of the Galactic plane.

catalogues are both too shallow to be used to reliably select the DESI survey targets. The DES survey (The Dark Energy Survey Collaboration 2005) does reach the target depth for DESI, but only covers 5000 deg^2, mostly in the South Galactic Cap (SGC), with only ∼1130 deg^2 observable with DESI.

This work is based on the eighth release of the Legacy Survey project (LS DR8) that is the first release to integrate data from all of the individual components of the Legacy Surveys (BASS, DECaLS, and MzLS). However, this paper focuses only on DECaLS data.

The DECaLS data in the LS DR8 data release comprise observations from 2014 August 9 through 2019 March 7. DECam images come from the Dark Energy Camera (DECam; Flaugher et al. 2015) at the 4-m Blanco Telescope at the Cerro Tololo Inter-American Observatory. DECam has 62 2048 × 4096 pixels format 250-μm-thick Lawrence Berkeley National Laboratory (LBNL) CCDs arranged in a roughly hexagonal ∼3.2 deg^2 field of view. The pixel scale is 0.262 arcsec pixel^{-1} and the camera has high sensitivity across a broad wavelength range of ∼400–1000 nm. Since LS DR8 data go beyond the intended DESI footprint[5] of ∼14 000 deg^2, we are going to consider only data within the DESI footprint. This corresponds to ∼9 717 deg^2 of DECaLS data of which ∼1 114 deg^2 are covered by DECam data coming from the DES (The Dark Energy Survey Collaboration 2005). We essentially have two DECam data sets: (i) DECam imaging taken for the LS programme that we refer to as DECam LS and (ii) the DECam data coming from the DES programme that we refer to as DECam DES. DECam LS and DECam DES combine to form the DECaLS data set. Fig. 1 shows the sky map coverage of DECaLS imaging indicating the DECaLS imaging that lies within the DESI footprint. DECaLS is the only survey that covers the entire SGC (4394 deg^2) and the North Galactic Cap (NGC; 5323 deg^2) regions of the DESI survey at declination $\delta \leq +32.^{\circ}375$.

In order to fulfil the target selection required for the different DESI surveys (BGS, LRGs, ELGs, and QSOs), it was concluded that a three-band *g*, *r*, and *z* optical imaging programme, complemented by *Wide-field Infrared Survey Explorer* (*WISE*) *W1* and *W2* photometry, would be sufficient. The minimal depth[6] required is $g = 24.0$, $r = 23.4$, and $z = 22.5$. DECam LS reaches these required depths in total exposure times of 140, 100, and 200 s in *g*, *r*, and *z*, respectively, in nominal[7] conditions, typically in a minimum of two visits per field.

All data from the Legacy Surveys are first processed at the National Science Foundation (NSF)'s National Optical-Infrared Astronomy Research Laboratory in Tucson (NSF's OIR Lab) through the NSF's OIR Lab Community Pipeline[8] (CP). The CP takes raw data as an input and provides detrended and calibrated data products such as instrumental calibration (e.g. bias subtraction and flat-fielding), astrometric calibration (e.g. mapping the distortions and providing a world coordinate system, or WCS), photometric characterization (e.g. magnitude zero-point calibration) and artefact identification, and masking and/or removal (e.g. removal of cross-talk and pupil ghosts, and identification and masking of cosmic rays).

The source catalogues for the Legacy Surveys are constructed using the LEGACYPIPE[9] software, which uses the TRACTOR[10] (Lang,

[5]Current LS DR8 imaging covers around ∼20 332 deg^2 of which 15 174 deg^2 corresponds to DECaLS.

[6]The depths are defined as the optimal-extraction (forced-photometry) depths for a galaxy near the limiting depth of DESI, where that galaxy is defined to be an exponential profile with a half-light radius of $r_{\rm half} = 0.45$ arcsec.

[7]Here 'nominal' is defined as photometric and clear skies with seeing FWHM of 1.3 arcsec, airmass of 1.0, and sky brightness in *g*, *r*, and *z* of 22.04, 20.91, and 18.46 AB mag arcsec^{-2}, respectively.

[8]https://www.noao.edu/noao/staff/fvaldes/CPDocPrelim/PL201_3.html

[9]https://github.com/legacysurvey/legacypipe

[10]https://github.com/dstndstn/tractor

Hogg & Mykytyn 2016) code for pixel-level forward-modelling of astronomical sources. This is a statistically rigorous approach to fitting the differing point spread functions (PSFs) and pixel sampling of these data, which is particularly important as the optical data have a typical PSFwidth of ∼1 arcsec.

The steps in the LEGACYPIPE processing are described in Dey et al. (2019); we briefly summarize relevant parts here.

After initial source detection and defining the contiguous set of pixels associated with each detection (termed a blob), LEGACYPIPE proceeds to fit these pixels with models of the surface brightness, including a point source and a variety of galaxy models. These fits are performed on the individual optical images (in *g*, *r*, and *z* bands), taking into account the different PSF and sensitivity of each image, using TRACTOR.

Besides the PSF model, TRACTOR fits four other light profile models to sources: a round exponential with a variable radius (referred to as REX), an exponential profile (EXP), a de Vaucouleurs profile (DEV), and a composite of DEV and EXP profiles (COMP). The decision as to whether or not to retain an object in the catalogue and the choice of the model to best describe its light profile are treated as a penalized-χ^2 model selection problem.

This process results in object fluxes and colours that are consistently measured across the wide-area imaging surveys that form the input into the DESI target selection. In general, TRACTOR improves the target selection for all DESI surveys by allowing information from low resolution and low signal-to-noise ratio measurements to be combined with those from high resolution and high signal-to-noise ratio data. The TRACTOR catalogues include source positions, fluxes, shape parameters, and morphological quantities that can be used to discriminate extended sources from point sources, together with errors on these quantities. The BGS is flux limited in the *r* band. However, since TRACTOR performs simultaneous fits in *g*, *r*, and *z* we also chose to impose quality cuts in the other bands and those in the *r* band when selecting the BGS targets.

The main TRACTOR outputs required for the BGS are the total fluxes[11] corresponding to the best-fitting source model (i.e. PSF, REX, EXP, DEV, or COMP) in all three bands (*g*, *r*, and *z*), the number of observations (NOBS) in the three bands, and the predicted flux (in the *r* band only) within the aperture of a fibre that is around 1.5 arcsec diameter (FIBERFLUX[12]) in 1 arcsec Gaussian seeing. The Galactic extinction values are derived from the SFD98 maps (Schlegel, Finkbeiner & Davis 1998) and are reported in linear units of transmission (MW_TRANSMISSION) in the *g*, *r*, and *z* bands, with a value of unity representing a fully transparent region of the Milky Way and zero indicating a fully opaque region. The extinction coefficients for the DECam filters were computed through an airmass of 1.3, for a source with a 7000 K thermal spectrum (Schlafly & Finkbeiner 2011). The resulting coefficients are $A/E(B-V) = 3.995, 3.214, 2.165, 1.592, 1.211$, and 1.064 in *ugrizY*. These are then multiplied by the SFD98 $E(B-V)$ values at the coordinates of each object to derive the *g*, *r*, and *z* MW_TRANSMISSION values. Finally, in each band, there is a set of quality measures called FRACMASKED, FRACFLUX, and FRACIN that quantify the quality of the data in each profile fit. We describe these in more detail in Section 4.4.

[11]The fluxes output by TRACTOR are in units called nanomaggies. A flux of 1 nanomaggie corresponds to an AB magnitude of 22.5.
[12]The FIBERFLUX is in units of nanomaggies.

Table 1. The area, in deg^2, of DECaLS DR8 covered by at least 1, 2, or 3 passes in each of the three filters (*grz*) individually (first three rows), and combined (i.e. at least 1, 2, or 3 passes in each of the three bands; bottom row). We have restricted our results to observations within the DESI footprint as shown in Fig. 1.

Band/Number of passes	≥1	≥2	≥3
g band	9687	9454	7769
r band	9686	9422	7569
z band	9686	9487	8036
Combined	9669	9257	6870

The fluxes returned by TRACTOR can be transformed into AB magnitudes as follows:

$$\mathrm{mag\,r} = 22.5 - 2.5\log_{10}(\mathrm{FLUX}), \quad (1)$$

$$\mathrm{mag} = 22.5 - 2.5\log_{10}(\mathrm{FLUX/MW\_TRANSMISSION}), \quad (2)$$

where equation (1) does not include the correction for Galactic extinction, unlike equation (2). The r in equation (1) stands for raw.

Table 1 shows the area covered by photometry in each of the three bands of DECaLS DR8 with 1, 2, or 3 passes. These values are just for the data within the DESI footprint, as shown in Fig. 1. This DECaLS footprint covers a total of 9717 deg^2. Expressed in percentages, 99.5 per cent of this area has at least one pass in all of the three bands *grz*, 95.3 per cent has at least two passes, and 70.7 per cent has at least three passes in all three bands.

2.2 Secondary catalogues

Here we list other catalogues that are used either to exclude regions of the sky in which the extraction of galactic sources is compromised by the presence of other objects, or to perform star–galaxy separation.

2.2.1 Tycho-2

Bright stars can impinge upon the estimation of the photometric properties of nearby galaxies or may even lead to the generation of spurious sources. Hence, it is prudent to simply exclude or veto regions close to known bright stars to avoid such problems. Regions near bright stars are masked out of the target catalogue using the Tycho-2 catalogue (Høg et al. 2000). The Tycho-2 catalogue contains positions, proper motions, and two-colour photometry for 2539 913 of the brightest stars in the Milky Way.

2.2.2 Gaia DR2

Gaia (Gaia Collaboration et al. 2016a) is a European Space Agency mission that was launched in 2013 with the aim of observing ≈1 per cent of all the stars in the Milky Way, measuring accurate positions for them along with their proper motions, radial velocities, and optical spectrophotometry. The wavelength coverage of the astrometric instrument, defined by the white-light photometric *G*-band magnitude, is 330–1050 nm (Carrasco et al. 2016). These photometric data have a high signal-to-noise ratio and are particularly suitable for variability studies.

Since the first release of *Gaia* data (Gaia Collaboration et al. 2016b), this survey has been widely used by the DESI LS (i.e. for astrometric calibrations, proper motions, bright star masking) and is also ideal for constructing a star–galaxy separator for the BGS.

There are 1.7 billion stars in the *Gaia* Data Release 2 (DR2),[13] over the whole sky to $G = 20.7$, which is sufficiently deep to detect all stars that might contaminate the BGS FAINT sample. We describe how we use a combination of *Gaia* and LS photometry to perform star–galaxy separation in Section 4.1.

2.2.3 Globular clusters and planetary nebulae

Globular clusters (GCs) and planetary nebulae are bright extended sources that can affect the identification of extragalactic sources in a similar way to bright stars. In the LS, an area of sky around such objects is excluded to minimize their impact on target selection. The OpenNGC catalogue[14] is used to provide a list of such sources. The extent and impact of masking around GCs and planetary nebulae is discussed in Section 3.1.3.

2.2.4 The Siena Galaxy Atlas

Large galaxy images can be broken up by photometric pipelines, which, for example, could mistake H II regions inside the galaxy for individual extended sources. Also, spurious sources could be generated around the boundaries of large galaxies. The Siena Galaxy Atlas – 2020 (SGA-2020)[15] is an ongoing project to select the largest galaxies in the LS using optical data from the HyperLeda catalogue[16] (Makarov et al. 2014) and infrared data from the AllWISE catalogue (Secrest et al. 2015). Currently the catalogue contains 535 292 galaxies that have an angular major axis (at the 25 mag arcsec^{-2} isophote) larger than 20 arcsec. The use of the SGA-2020 in the spatial mask of the BGS is described in Section 3.1.2.

3 SPATIAL MASKING

Our main goal is to produce a reliable BGS input catalogue that fulfils the DESI science requirements. If the target list contains spurious objects, these will mistakenly be allocated fibres leading to a reduction in the efficiency and completeness of the redshift survey. Furthermore, spurious objects could imprint a systematic effect in the measured clustering.

A step towards minimizing the number of spurious objects is to mask out regions of the sky around bright stars, since features such as extended haloes, ghosts, bleed trails, and diffraction spikes around the stars can compromise the measurement of the photometry of neighbouring objects. Similarly we must remove areas around very large galaxies and GCs and planetary nebulae; such objects can also affect the photometric measurements of their neighbours, leading to incorrect properties or spurious objects.

Within the same framework, we have to propagate instrumental effects such as saturated pixels, bad pixels, bleed trails, etc. that the NSF's OIR Lab CP tracks and TRACTOR reports in the LS catalogue.[17]

One way to avoid contamination of the catalogue with spurious objects is to exclude regions around bright stars and galaxies. This can be done with a simple but effective circular mask for stars and by using elliptical masks for galaxies. In Section 3.1, we set out the geometrical masking functions we have applied around bright stars, large galaxies, and GCs to minimize the number of spurious targets in our BGS catalogue. In Section 3.2, we describe the masks applied to reduce the number of spurious targets due to imaging artefacts such as bad pixels resulting from saturation and bleed trails.

For subsequent analysis (e.g. estimating clustering statistics), it is very important to keep a record of the areas of the survey that are removed by these masks. For this purpose we have made use of the random catalogue developed by the DESITARGET[18] team. The random catalogue has a total density of 50 000 objects deg^{-2} divided into 10 subsets, each with density of 5000 objects deg^{-2}. Each random carries with it some of the DEC amimaging information computed from the image pixel (in each band and exposure) in which it is located and supplementary information such as the dust extinction extracted from HEALPIX[19] maps (Zonca et al. 2019). These imaging attributes include the number of observations (NOBS_G, NOBS_R, NOBS_Z), galactic extinction (EBV), the bitwise mask for optical data (MASKBITS), etc.[20]

In Fig. 2, we show a flow chart that summarizes the spatial masking applied when constructing the BGS catalogue. The spatial masking is broken down into two classes: *geometrical masking* and *pixel masking*. The blue boxes of the flow chart report the survey area (in deg^2) and mean target densities (in objects deg^{-2}) after successively applying each mask (grey hexagonal boxes). The red boxes record the same information for the rejected area and objects. The final BGS catalogue does not depend on the order in which the masks are applied, but as some areas and targets are rejected by more than one mask the information in the red boxes depends on the ordering. For example, the area and number of objects shown as being rejected by the pixel masking excludes what would be rejected by this mask if the geometric masks had not been applied first. Overall, for the DECaLS footprint of 9717 deg^2, the spatial masking removes 3.25 per cent of the area.

3.1 Geometrical masking

3.1.1 Bright star mask

The bright star (BS) mask is based on the locations of stars from *Gaia* DR2 (Gaia Collaboration et al. 2018) and the Tycho-2 (Høg et al. 2000) catalogue after correcting for epoch and proper motions. This mask consists of the union of circular exclusion regions around each star, where the radius of the exclusion region, estimated from an earlier stacking analysis, depends on the magnitude of the star in the following way:

$$\begin{aligned} R_{\rm BS}(m) &= 39.3 \times 2.5^{(11-m)/3}\,\text{arcsec}, & m > 2.9 \\ &= 471.6\,\text{arcsec}, & m < 2.9. \end{aligned} \quad (3)$$

Here m is either *Gaia* G-mag or Tycho-2 MAG_VT with *Gaia* G-mag being used when both are available. Stars fainter than $m = 13$ have no exclusion zone around them.

The BS masking uses a total of 773 673 *Gaia* DR2 objects (82 objects deg^{-2}) with *Gaia* G-mag brighter than 13, while from Tycho-2, we have a total of 3349 objects (∼0.36 objects deg^{-2}) to a Tycho-2 visual magnitude brighter than MAG_VT =13. In order to

[13] DR2 covers 22 months of observations and was released on 2018 April 25.
[14] https://github.com/mattiaverga/OpenNGC
[15] https://github.com/moustakas/SGA
[16] http://leda.univ-lyon1.fr/
[17] In the LS DR8 catalogue information on whether or not the photometric parameters measured for an object have the possibility of being influenced by a bad pixel is flagged by the ALLMASK MASKBITS.
[18] https://github.com/desihub/desitarget
[19] http://healpix.sourceforge.net
[20] For more information on the properties of randoms see: http://legacysurvey.org/dr8/files/#random-catalogs

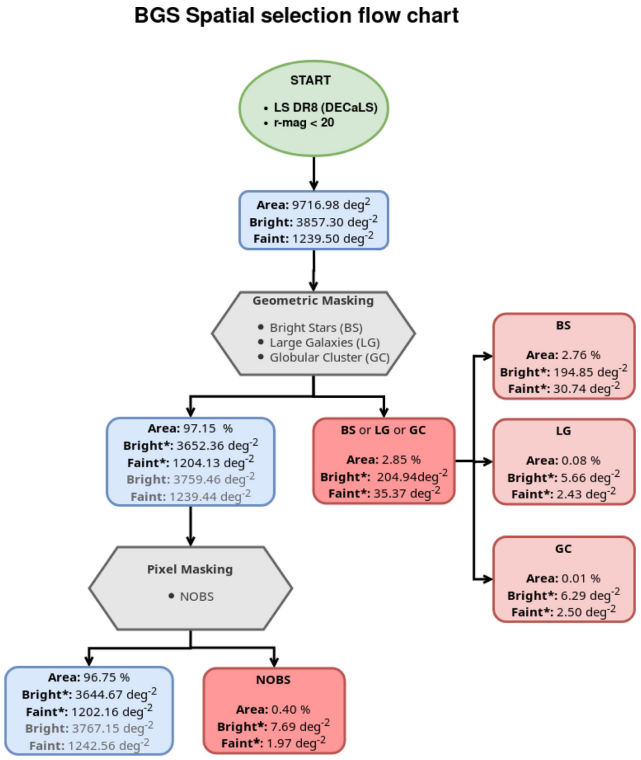

Figure 2. The flow chart shows the effects of the spatial masks that are applied as part of BGS target selection for the DECaLS DR8 data. The spatial masking is divided into two classes: one defined by the geometrical cuts that exclude regions around bright sources (bright stars, large galaxies, and globular clusters), and the other by pixel-based cuts that use information such as the number of observations (NOBS). The boxes in the flow chart show the survey area (in deg^2) and the target number density (per deg^2) split into BGS BRIGHT and BGS FAINT after each mask is applied. The blue boxes give this information for the portion of the survey that is retained, while the red boxes give this information for the areas removed. If more than one mask is combined at a single stage (as indicated within the grey hexagonal boxes), then the dark-red boxes show the results for the combination of these masks and the light-red boxes show the results for each individual mask. As some of the masks can overlap the numbers in the light-red boxes do not necessarily add up to those in the dark-red boxes. The target densities with the superscript asterisk (*) are computed without correcting for the area removed by the masking, while those without the superscript asterisk (*) are corrected for the masked area. The grey hexagonal boxes describe the different masks. Note that star–galaxy separation is not yet applied here and this is why we have a high target density in the blue boxes.

avoid overlaps both catalogues have been matched after applying proper motions to bring *Gaia* objects to the same epoch as Tycho-2 and keeping only the Tycho-2 objects that are not found in *Gaia*. These Tycho-2 stars represent only a 0.4 per cent of total stars used for the BS masking. Then the magnitude, m, used to compute the mask radius in equation (3) is the *Gaia* G-band magnitude for the *Gaia* stars and the Tycho-2 visual magnitude, MAG_VT, for the retained Tycho-2 stars. The overall median difference between the Tycho-2 and *Gaia* magnitude is 0.4 with Tycho-2 being fainter. This 0.4 mag difference translates into a median decrease in masking radius of 50 arcsec for *Gaia* stars with magnitude of 3 and a decrease of 2 arcsec for *Gaia* stars with magnitude of 13 from equation (3). Within $R_{\rm BS}(m)$ TRACTOR forces all the sources it detects to be fit with the PSF profile to avoid artificially fitting diffraction spikes and stellar haloes as large extended sources. Thus any galaxies detected within $R_{\rm BS}$ will have their fluxes underestimated. Consequently to define a reliable galaxy catalogue we must veto all sources within $R_{\rm BS}$ of a bright star. In Fig. 2, we show that this bright star mask covers 2.76 per cent of the initial footprint and rejects ∼195 potential BGS BRIGHT objects deg^{-2} and ∼31 potential BGS FAINT objects deg^{-2} when averaged over the full initial footprint. It should be noted that most of these objects are stars as star–galaxy separation has not been applied at this stage in the flow chart shown in Fig. 2. An alternative ordering of the flow chart with star–galaxy separation applied first is shown in Fig. A1. There we see that for galaxies the corresponding numbers are 13.7 galaxies deg^{-2} for BGS BRIGHT and 8.5 galaxies deg^{-2} for BGS FAINT.

To determine if the bright star mask is adequate or whether the effects of stellar haloes cause a systematic error in the photometry of neighbouring galaxies that extend to larger radii, we plot in Fig. 3 the average density of BGS galaxies in the vicinity of bright stars prior to applying the bright star mask. If the photometry of galaxies has been compromised in any means, this can be seen in the galaxy number density to a fixed magnitude due to the strong dependence of galaxy number density on apparent magnitude. The term BGS galaxy refers to the BGS sample after applying the star–galaxy separation and the spatial and photometric cuts down to the r-band magnitude of 20, which will be covered in the subsequent subsections of Sections 3 and 4. The stacks are made by expressing the angular separation, r, of the BGS galaxies prior to apply the bright star mask from their nearest bright star in units of the bright star masking radius $R_{\rm BS}$, as given by equation (3). In these rescaled coordinates, $R = r/R_{\rm BS}$, galaxies within a radius of unity, shown by the black circle, are within the BS masking zone. We show stacks for two magnitude bins defined by the G-mag and visual magnitude MAG_VT for *Gaia* DR2and Tycho-2 stars, respectively, one with bright stars of magnitude between 8 and 12 and one fainter with magnitude between 12 and 13. The radial profile (red solid line) shows the variation in the target density, defined as $\Delta\rho(R) \equiv \log_2(\eta(R)/\bar{\eta})$ where $\eta(R)$ is the target density in an annulus at radius R of width $\Delta R \sim 0.06$, and $\bar{\eta}$ is the mean target density evaluated over the region $1.1 < R < 3$. This means that $\Delta\rho(R) = 0$ corresponds to the mean density, $\Delta\rho(R) \geq 1$ to an overdensity at least twice the mean density, and $\Delta\rho(R) < 0$ to an underdensity. The large underdensity at radius $R \leq 1$ is due to TRACTOR forcing all objects within this region to be fit by the PSF model. In Section 4.1, we will see how stars and galaxies are defined for BGS target selection, which does not depend on TRACTOR PSF designation, therefore, galaxies in the region $R < 1$ are allowed. In the left-hand panel of Fig. 3, we see a spike of spurious galaxies for $R < 0.2$. In contrast the right-hand panel shows a strong deficit of galaxies at $R < 0.2$. For $R > 1$, the stacks show uniform density close to mean, suggesting the star mask is working. There is a small bump just outside the masking radius where a ∼6 per cent excess is seen in both panels. This may need to be revisited for accurate clustering studies, but is not large enough to be a concern for the efficiency of target selection.

3.1.2 *Large galaxies mask*

Without special treatment, large galaxies in which spiral arms and other structures such as H II regions are resolved would be artificially fragmented by TRACTOR into multiple sources. To avoid this and to achieve more accurate photometry for large galaxies in the SGA-2020 catalogue (see Section 2.2.4), TRACTOR is seeded with different priors, and within an elliptical mask centred on the large galaxy TRACTOR fits secondary detections using only the PSF

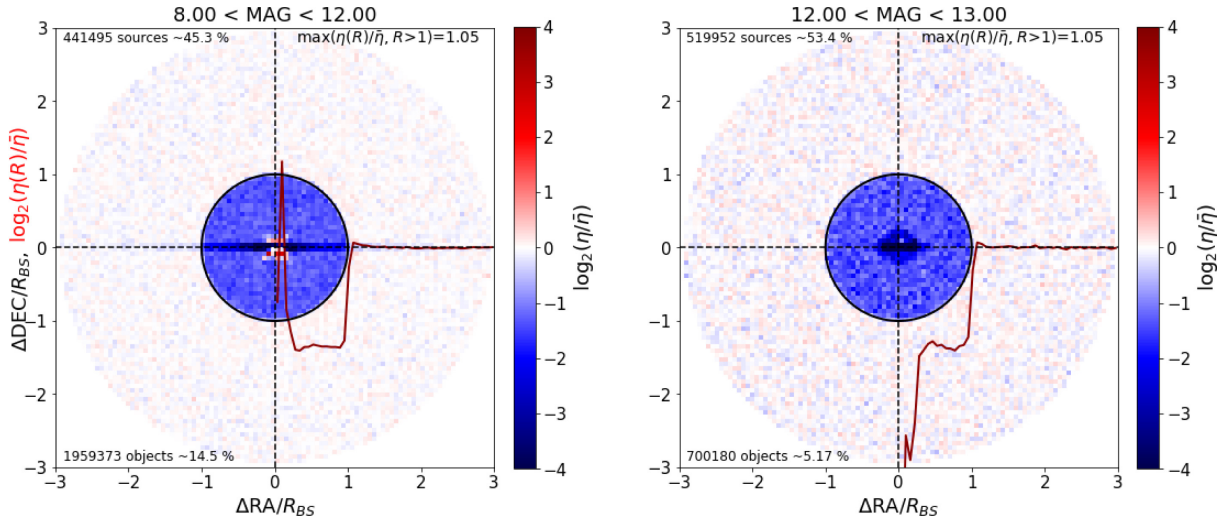

Figure 3. 2D histograms of the positions of BGS objects relative to their nearest Bright Star (BS) taken from the *Gaia* and Tycho-2 sources down to *G*-mag and visual magnitude MAG_VT of 13, respectively. These stacks are performed in magnitude bins in the BS catalogue from magnitude 8 to 12 (left) and from 12 to 13 (right). The stacks are made using angular separations rescaled to the masking radius function given in equation (3), which means that objects within a scaled radius of 0 to 1 will be masked out by the BS veto while objects with $R = r/R_{\rm BS} > 1$ will not (here $r^2 = (\Delta{\rm RA}^2\cos({\rm Dec.})^2 + \Delta{\rm Dec.}^2)$. The colour scale shows the ratio of the density per pixel (η) to the mean density ($\bar{\eta}$) within the shell $1.1 < r/R_{\rm BS} < 3$. The density ratio is shown on a $\log_2$ scale where red shows overdensities, blue corresponds to underdensities, and white shows the mean density. The black solid circle shows extent of the BS exclusion zone. The red solid line shows the radial density profile on the same scale as the colour distribution $\log_2(\eta(R)/\bar{\eta})$, where $\eta(R)$ is the target density within the annulus at radius R of width $\Delta R \sim 0.06$.

model. This reduces the spurious fragmentation of large galaxy images, but also means that genuine neighbouring galaxies within the masked area have compromised photometry. The elliptical mask that is used has the same position, 25 mag arcsec^{-2} isophotal major axis angular diameter, D_{25}, semiminor to semimajor ratio, B/A, and position angle, PA, as the ones used to define the large galaxies in the SGA-2020 catalogue. Defining an effective masking radius of $r = \sqrt{ab}$, where a and b are the semimajor and semiminor axes of the elliptical mask, the median masking radius for the LG galaxies is 10.8 arcsec.

We apply these same masks to reject objects from the BGS catalogue, but then we reinstate the large galaxies provided they are not also masked by the bright star or GC mask. The area covered by the combined LG mask amounts to only 0.08 per cent of the initial area and the number of objects removed amounts to 5.7 objects deg^{-2} BGS BRIGHT and 2.4 objects deg^{-2} BGS FAINT objects over the full initial area.

3.1.3 Globular cluster mask

The globular cluster (GC) mask works in a similar way to the BS mask, by applying a circular exclusion zone around the GC. The masking radius is defined by the major axis attribute for the object in the OpenNGC catalogue.

The GC mask has the smallest impact of the geometric masks, rejecting only 0.01 per cent of the initial area, accounting for densities of 6.3 objects deg^{-2} in BGS BRIGHT and 2.5 objects deg^{-2} in BGS FAINT. TRACTOR also force fits as PSFs everything within this mask.

3.2 Pixel masking

Some of the effects that compromise the photometry on a pixel basis and the model fitting include bad pixels, saturation, cosmic rays, bleed trails, and transients. The NSF's OIR Lab DECam CP identifies these instrumental effects during its various calibrations[21] (see table 5 in Dey et al. 2019 for a list of the calibrations) and these are passed through TRACTOR and compiled in the ALLMASK BITMASK.[22] ALLMASK denotes a source blob that overlaps with any of the mentioned bad pixels in all of the overlapping images.

Besides the bad pixels that arise due to instrumental defects, the BGS requires a complete sample in the three bands (g, r, and z). We therefore impose a requirement that there is at least one observation in each of the bands through the NOBS parameter. NOBS stands for number of observations, and is defined as the number of images that contributes to the source detected central pixel in each of the bands. Both ALLMASK and NOBS are pixel based and hence this information is also available in the random catalogue. However, we find that virtually all of the area (97 per cent) (and hence virtually all of the randoms) rejected by ALLMASK is also rejected by using NOBS=0 (in any band). In addition, ALLMASK rejects a significant number of objects (196 objects deg^{-2}) but with a small associated area (0.01 per cent of the full area). Virtually all the objects rejected by ALLMASK and many others are already rejected by the quality cuts in FRACMASKED, FRACIN, and FRACFLUX (in any band); these cuts will be reviewed in Section 4.

In conclusion, there is little to be gained from using ALLMASK and we have therefore decided to use only NOBS as our pixel level mask, shrinking the area by 0.4 per cent and reducing the target density by 7.7 objects deg^{-2} in BGS BRIGHT and 2 objects deg^{-2} in BGS FAINT.

[21]The document that lists all the calibrations and that includes details about the various maskings can be found at: https://www.noao.edu/noao/staff/fvaldes/CPDocPrelim/PL201_3.html

[22]Details of this BITMASK can be found here: http://www.legacysurvey.org/dr8/bitmasks/#allmask-x-anymask-x

4 PHOTOMETRIC SELECTION

Following the spatial masking described in the previous section, the next step in the construction of the BGS target list is to incorporate information about photometric measurements into the selection process. According to the science requirements of the BGS and the mock BGS catalogues made by Smith et al. (2017), the survey is expected to have a target density of 800 galaxies deg^{-2} to an r-band limit of 19.5. For the faint sample ($19.5 < r < 20$), which is second priority in BGS, a density of 600 galaxies deg^{-2} is expected.

One of the major challenges for the BGS is the separation of stars and galaxies. In Section 4.1, we describe how we compare high angular resolution point source magnitudes from *Gaia* DR2 (Gaia Collaboration et al. 2018) with total magnitudes from the best-fitting light profile model selected by TRACTOR to distinguish point sources from extended sources.

In Section 4.2, we describe how we reject spurious objects that have incongruous light profiles by comparing their total magnitudes with the fibre magnitude that TRACTOR computes from the fitted profile assuming 1 arcsec Gaussian seeing and 1.5 arcsec fibre diameter. We place a cut in the fibre magnitude versus total magnitude plane that is motivated by the locus of confirmed galaxies from the GAMA DR4 survey.

Further posterior cuts that use photometry include removing colour outliers in $g - r$ and $r - z$ (see Section 4.3), and applying quality cuts that indicate low accuracy in the flux measurement for an object (see Section 4.4). The quality cuts make use of the quantities FRACMASKED, FRACFLUX, and FRACIN measured by TRACTOR for each object in each of the three bands (grz). These are defined and discussed in Section 4.4.

In Fig. 4, we show the second part of the BGS target selection flow chart. This flow chart focuses on the photometric selection cuts and starts from where the previous flow chart (Fig. 2), showing the spatial cuts, left off. The BGS catalogue, in the DECaLS subregion, ends up having a reduced area of 9401 deg^2 out of the initial 9717 deg^2, and target densities of 846 and 578 objects deg^{-2} for BGS BRIGHTand BGS FAINT, respectively.

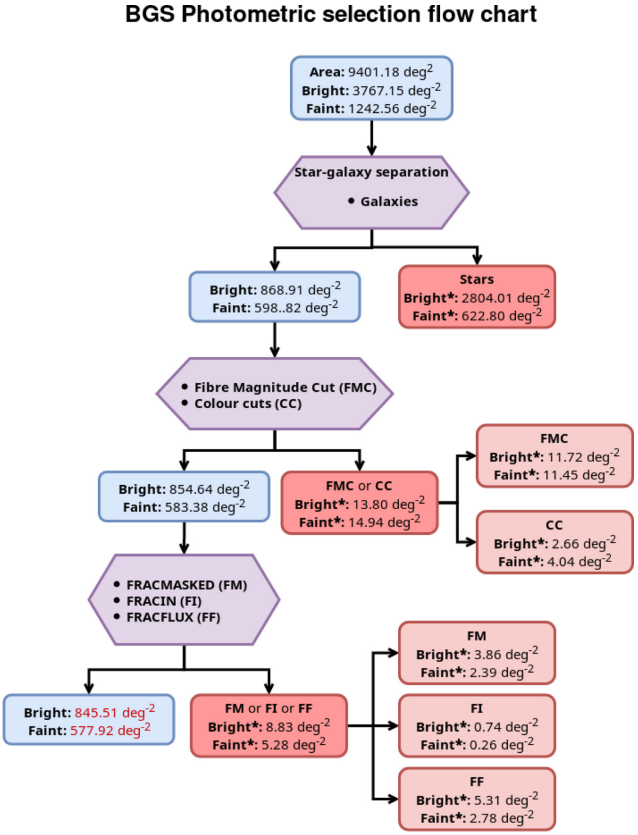

Figure 4. Flow chart of the BGS target selection in the Legacy Surveys DR8 based on photometric considerations. The photometric selection of BGS targets is divided into four stages: star–galaxy separation; fibre magnitude cuts (FMC); colour cuts (CCs); and quality cuts (QCs). The photometric cut flow chart is a continuation of the spatial cut flow chart (Fig. 2) and therefore we start from the area and object densities reported at the end of the spatial cut flow chart. We report densities for the bright and faint samples separately, showing in blue boxes the values for the sources remaining after each of the BGS cuts. The densities of the removed objects are shown in red/pink boxes. The different cuts applied are shown in purple hexagonal boxes.

4.1 Star–galaxy separation

The classification of images as star or galaxies is an old problem that is of great importance when defining target catalogues for the efficient use of multi-object spectrographs. Sophisticated techniques are employed that include algorithms using machine learning methods applied to both colour and morphological information, e.g. artificial neural networks (Odewahn et al. 1992; Bertin & Arnouts 1996), support vector machines (Fadely, Hogg & Willman 2012), and decision trees (Weir, Fayyad & Djorgovski 1995). TRACTOR uses a rigorous statistical approach to determine the best-fitting light profile model to each object. In this way it classifies objects as either point sources (PSF) or extended sources (DEV, EXP, COMP, or REX). However, this pipeline is not infallible and it is inevitable with ground-based seeing that some compact galaxies will be misclassified as being of PSF type rather than extended. As we want to avoid incompleteness that depends on the variable seeing of the images, we have instead made use of the space-based high angular resolution *Gaia* photometry to distinguish point sources from extended sources. This is possible for the BGS as virtually[23] all stars brighter than the BGS magnitude limit of $r < 20$ are bright enough to be detected by *Gaia*.

The *Gaia* DR2 catalogue (Gaia Collaboration et al. 2018) that we use is primarily a catalogue of stars but has some galaxy and quasar contamination as reported by Bailer-Jones, Fouesneau & Andrae (2019). This means we cannot simply classify all of the BGS objects that are in *Gaia* as stars. However, by comparing TRACTOR magnitude measurements with the higher spatial resolution magnitude measurements from *Gaia* we can determine which objects have extended light profiles. The *Gaia* magnitudes are computed assuming all objects are point sources. This results in accurate magnitudes for stars but magnitudes that are systematically fainter than the associated total magnitudes for sources that are extended compared to the ∼0.4 arcsec PSF achieved by *Gaia*. In contrast, the model magnitudes computed by TRACTOR should capture more fully the total magnitude of the object. Consequently, if *Gaia* and TRACTOR magnitudes were measured in the same band, we would expect them to agree for point sources but for the TRACTOR magnitude to be brighter than the *Gaia* magnitude for extended sources. We would even expect this to be true for extended objects that TRACTOR misclassifies as PSF since the wide, ground-based PSF of TRACTOR would capture more of the total flux than the narrow PSF of *Gaia*.

[23] *Gaia* DR2 is complete in the range $12 < G < 17$ mag.

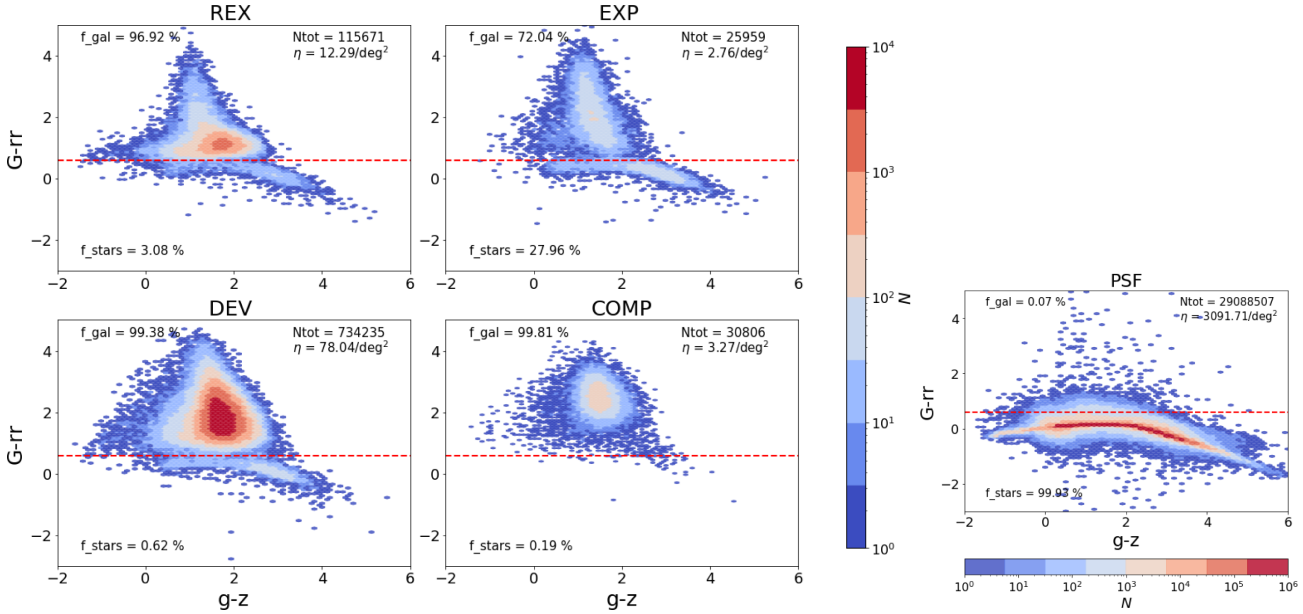

Figure 5. Separately for objects classified by TRACTOR as type REX, EXP, DEV, COMP, and PSF, we show the difference between the *Gaia* (PSF) magnitude G and total non-dust corrected r-band model magnitude measured by TRACTOR, rr versus TRACTOR extinction corrected $g - z$ colour. All the objects plotted have passed the geometrical and pixel cuts detailed in Fig. 2, and all but the star–galaxy classification cut of the photometric-based cuts detailed in Fig. 4. The plots show objects that have been cross-matched between LS DR8 objects and *Gaia* DR2. Each panel shows a different morphological class, as labelled, according to the best-fitting light profile assigned by TRACTOR. The red-dashed line indicates our adopted division at $G - \mathrm{rr} = 0.6$ with stars below and galaxies above the line. The colour in the plots shows the number counts of objects in a hexagonal cell, ranging from 1 to 10 000, except for the case of PSF-type objects, in which case the colour scale covers the range from 1 to 1 million as indicated in the colour bars. We display the fraction of galaxies and stars according to this classification at the top left-hand corner and bottom left-hand corner, respectively. The total number of objects (N_{tot}) in each plot and the target density (η) this represents is displayed in the top right-hand corner.

The complication is that the *Gaia* G band is a much wider filter than the DESI r band, but as we shall see, the colour dependence is weak.

Based on these considerations we define TRACTOR objects with $r < 20$ as being galaxies if either of the following two conditions is met.

(i) The object is not in the *Gaia* catalogue.
(ii) The object is in the *Gaia* catalogue but has $G - \mathrm{rr} > 0.6$.

In the above, the G band is the G photometric *Gaia* magnitude and rr is the raw r-band magnitude from the LS DR8 *without* applying a correction for Galactic extinction. This choice is made because the *Gaia* magnitude is not corrected for Galactic extinction. The discussion above explains that G and rr magnitudes are measured in different effective apertures and so the quantity $G - \mathrm{rr}$ should be thought of as a measure of how spatially extended an object is and not its colour. The first criterion above is satisfied by most (93 per cent) of the BGS objects. It leaves very little stellar contamination in the BGS, as essentially any star brighter than $r = 20$ is bright enough to be detected and catalogued by *Gaia*. The second criterion is required to keep the BGS completeness high by not rejecting galaxies that are in the *Gaia* catalogue.

In Fig. 5, we show the $G - \mathrm{rr}$ versus $g - z$ plane for objects in *Gaia* DR2 that are matched with objects in the LS DR8. The panels show different objects as classified by the TRACTOR model fits (i.e. PSF, COMP, DEV, EXP, and REX). The cross-matched objects have been subject to all the BGS cuts (i.e. both spatial and photometric) with the exception of the star–galaxy separation itself. For objects classified by TRACTOR as PSF type, we can see the stellar locus around $G - \mathrm{rr} = 0$ with a weak colour dependence. For the extended sources (i.e. COMP, DEV, EXP, and REX), we see part of the galaxy locus[24] in the upper part of the plot, just above $G - \mathrm{rr} = 0$.

From Fig. 5 we can see that the assignment of the best-fitting TRACTOR model supports our *Gaia* classification using $G - \mathrm{rr} > 0.6$, but we can still see some remnants of the stellar locus for objects that have not been assigned PSF type by TRACTOR. For the objects classified PSF type by TRACTOR we see in the rightmost panel of Fig. 5 that 99.93 per cent fall on the stellar side of our $G - \mathrm{rr}$ cut. For the objects classified by TRACTOR as the extended types (REX, DEV, and COMP) the stellar contamination (i.e. objects with $G - \mathrm{rr} < 0.6$) is at most 3.1 per cent. However, the contamination of the EXP-type objects is approaching 30 per cent.

The BGS target selection has the expected surface density after applying the star–galaxy separation. From the spatial cut flow chart in Fig. 4, we find a bright target density of 868.91 objects deg^{-2} and a faint target density of 598.82 objects deg^{-2}. Rejected *Gaia* stars have a target density of 2804.01 objects deg^{-2} bright stars and 622.80 objects deg^{-2} faint stars.

4.2 Fibre magnitude cut

In order to reduce the number of image artefacts and fragments of 'shredded' galaxies that would otherwise be classified as BGS targets, we apply a cut on the fibre magnitude that is defined as a

[24]We have to remember that Fig. 5 only includes stars and galaxies that are cross-matched between LS DR8 and *Gaia* DR2.

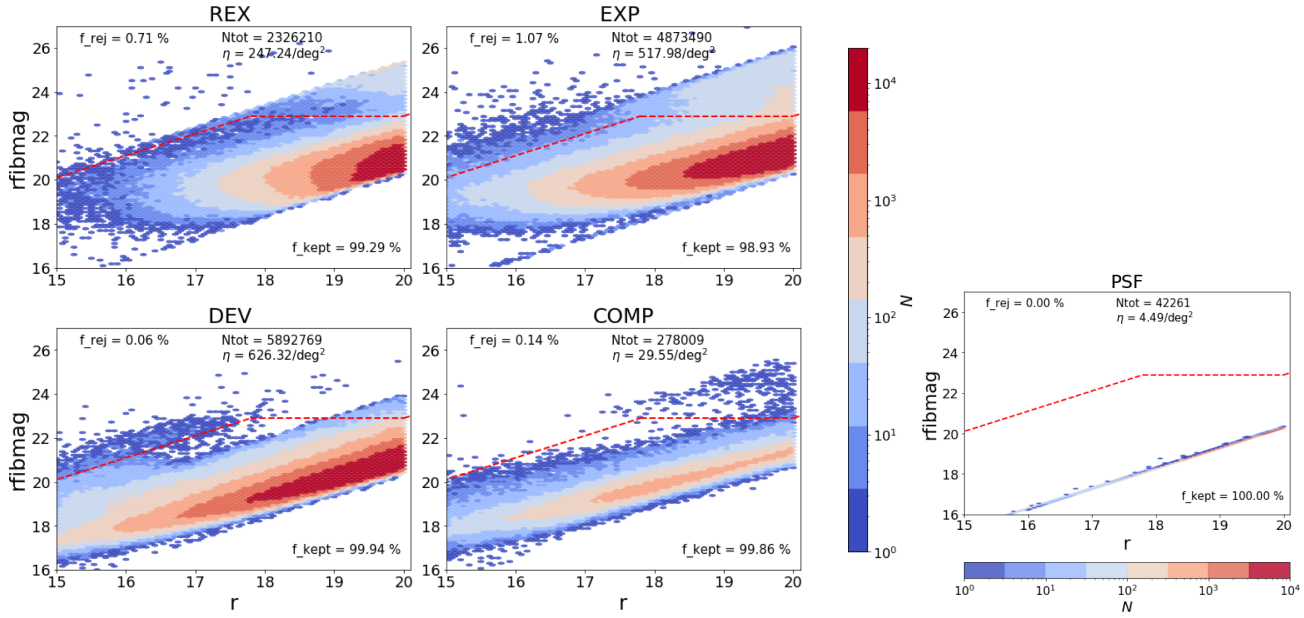

Figure 6. BGS galaxies in the *r*-band total magnitude (*x*-axis) versus *r*-band fibre magnitude (*y*-axis) plane in the LS DR8. The results are divided into the five different TRACTOR best-fitting light profile models, as labelled at the top of each panel. The colour bar shows the number counts of objects in a hexagonal cell covering the range from 1 to 20 000 for four of the light profile models with the exception of PSF-type galaxies, in which case the scale covers 1 to 10 000. The red-dashed line shows the fibre magnitude cut (FMC): we reject every object that is above this threshold. The numbers shown in top left-hand and bottom right-hand corners give the fraction of galaxies rejected and kept, respectively, while the number in the top right-hand corner shows the total number of galaxies (Ntot) and the corresponding target density (η).

function of *r*-band magnitude as follows:

$$\text{rfibmag} < \begin{cases} 22.9 + (r - 17.8) & \text{for } r < 17.8, \\ 22.9 & \text{for } 17.8 < r < 20, \end{cases} \quad (4)$$

where rfibmag is the magnitude of the predicted *r*-band fibre flux and *r* is the total *r*-band magnitude, both extinction corrected. The location of this cut was guided by inspecting postage stamp images of a selection of the objects with the faintest fibre magnitudes with the aim of rejecting objects that appear to be artefacts while retaining nearly all of the genuine galaxies. In addition, at the bright end our threshold was guided by the location of spectroscopically confirmed GAMA galaxies, as discussed further in Section 5.1. Fig. 6 shows the distribution of the BGS objects in the rfibmag versus rmag plane, with a separate panel for the different TRACTOR classes, and a red-dashed line indicating the location of the *fibre magnitude cut* (hereafter FMC). In the first four panels, we can see that the galaxy locus has a tight core and, in general, is well below the FMC. The FMC removes 1.2 per cent of the objects classified as EXP and even smaller fractions of the other light profile classes.

All BGS objects in the PSF class lie on a stellar locus. Whether all these objects are stars or whether this is an artefact of TRACTOR-only fitting the PSF model to *Gaia* sources with low astrometric excess noise (AEN) is revisited in Section 5.1, where we compare our classification with that of the GAMA DR4 survey. The stellar locus is also visible in the other photometric classes indicating there is some stellar contamination in our sample, but it is at a very low level.

In summary the adopted FMC rejects a further 23.17 objects deg^{-2} of which 11.72 are in BGS BRIGHT and 11.45 are in BGS FAINT from the objects that have passed the previous cuts that include the rejection of stars by our star–galaxy classifier.

4.3 Colour cuts

An efficient way of rejecting further spurious targets from the BGS is to reject objects with bizarre colours. The limits we impose to reject outliers are

$$-1 < g - r < 4,$$
$$-1 < r - z < 4. \quad (5)$$

Fig. 7 shows the $g - r$ versus $r - z$ colour–colour distribution of the objects retained in BGS if all but the colour cuts (CCs) were applied. The red box indicates the colour range we keep. We can see from this plot that the locus of normal galaxy colours lies well within the range we retain and the cuts are only removing objects/artefacts with bizarre colours. It is evident that some stellar contamination remains as the stellar locus can be seen as a spur of objects with very red $r - z$ colours. However the density of objects in this spur, and its blueward extension that overlaps the galaxy locus, is no more than a few objects deg^{-2} as we shall see in Section 5.1. The CCs we apply reject an additional 6.7 objects deg^{-2}, with 2.66 in BGS BRIGHT and 4.04 in BGS FAINT.

4.4 Quality cuts

Each object in the TRACTOR catalogue has three measures of the quality of its photometry recorded in each of the three bands (*grz*). These are as follows.

(i) FRACKMASK (FM). The profile-weighted fraction of pixels masked in all observations of the object in a particular band. This quantity lies in the range [0,1]. High values indicate that most of the flux of the fitted model lies in pixels for which there are no data due to masking and so the measurement is unreliable.

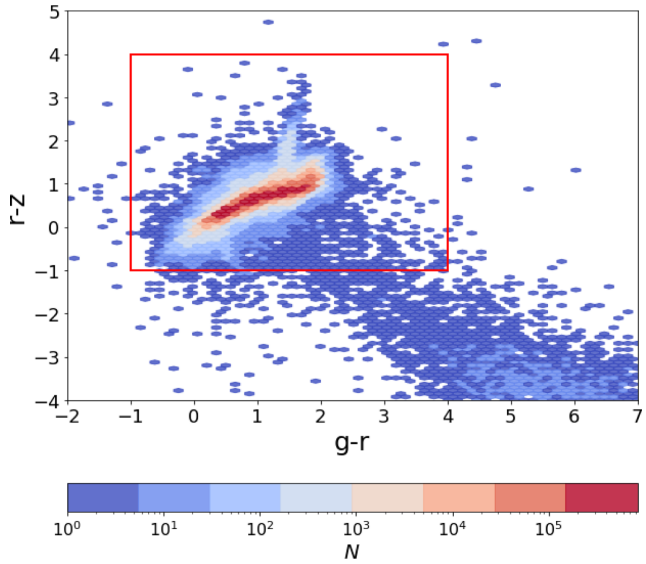

Figure 7. Colour–colour distribution showing $g - r$ versus $r - z$ for BGS objects without applying the CCs. The colour bar shows the number counts of objects in a hexagonal cell covering the range from 1 to 800 000. The solid red box shows CCs defined in equation (5). Sources outside of this box are excluded from the BGS.

(ii) FRACIN (FI). The fraction of the model flux that lies within the set of contiguous pixels (termed a 'blob') to which the model was fitted. FRACIN is close to unity for most real sources. Low values indicate that most of the model flux is an extrapolation of the model into regions in which no data were available to constrain it.

(iii) FRACFLUX (FF). The profile-weighted fraction of the flux from other sources divided by the total flux of the object in question. FRACFLUX is zero for isolated objects but can become large for faint objects detected in the wings of brighter objects that are nearby.

Once the other cuts have been applied, in particular, the cut on NOBS and the BS mask, the distribution of each of these quantities is tightly peaked around the favoured values of FRACMASKED ≈ 0, FRACIN ≈ 1, and FRACFLUX ≈ 0. However, each quantity has a distribution with a fairly featureless tail that extends out to less desirable values. There are also clear correlations between the three quantities for a given photometric band and in some cases between photometric bands. The choice of the best set of thresholds to reject outliers is not trivial. We have adopted the following quality cuts (QCs):

$$\begin{aligned} \text{FRACMASKED\_}i &< 0.4, \\ \text{FRACIN\_}i &> 0.3, \\ \text{FRACFLUX\_}i &< 5, \qquad \text{where } i = g,\ r,\ \text{or } z, \end{aligned} \tag{6}$$

based on visual inspection of postage stamp images.

As mentioned in Section 3.2, we find that the objects flagged by the TRACTOR quantity ALLMASK are essentially a subset of the objects that are rejected by applying the quality cuts listed in equation (6). While cutting on ALLMASK would have the advantage that it could also be applied to the randoms, we find that it is important to apply the QCs to remove spurious objects that are missed by the other cuts. For instance, some spurious objects that are outliers in either the fibermag versus mag plane or in the colour–colour space that just pass the FMC and CCs are removed by considering FRACMASKED or FRACIN.

Table 2. The BGS target densities for each of the TRACTOR best-fitting photometric models. The first column labels the photometric model. The next three columns list the surface density of objects per deg^2 for the BGS BRIGHT and BGS FAINT samples separately and their combined sum. The area covered by the DECaLS portion of the BGS is 9401 deg^2.

Model	$\eta_{\rm bright}$ (deg^{-2})	$\eta_{\rm faint}$ (deg^{-2})	$\eta_{\rm overall}$ (deg^{-2})
DEV	427	202	629
EXP	284	230	514
REX	104	141	246
COMP	27	3	31
PSF	3	2	5
Total	846	578	1423

As shown in the flow chart, Fig. 4, the QCs reject an additional 14.11 objects deg^{-2} of which ∼60 per cent are removed by FRACFLUX, ∼45 per cent by FRACMASKED, and ∼7 per cent due to FRACIN. The overlap between the FRACMASKED, FRACIN, and FRACFLUX cuts is minimal, with only 1.05 objects deg^{-2} for objects with $r < 19.5$, and in round 0.15 objects deg^{-2} for objects with $19.5 < r < 20$ being rejected by more than one of the cuts. Separately for BGS BRIGHT and BGS FAINT, we show the target density of objects rejected by these cuts after applying all the previous cuts. The largest overlap between these cuts is between FRACMASKED and FRACFLUX for BGS BRIGHT, but even here it amounts to less than 1 object deg^{-2}. For BGS FAINT this overlap is small, 0.11 object deg^{-2}, and there is no overlap with FRACIN.

In Appendix A, we present another version of the selection cut flow chart in which the cuts are applied in a different order. There we give a galaxy view of the target selection by first applying the star–galaxy classification so that all the subsequent cuts apply only to galaxies. The final selected sample that comprises of 845.5 galaxies deg^{-2} in BGS BRIGHT and 577.9 galaxies deg^{-2} in BGS FAINT is exactly the same, as the order of the cuts does not matter. The objects rejected by each filter, however, do change as many objects are rejected by more than one filter. To illustrate this point we have also swapped the order of the FMC and QCs cuts so one can see how these influence one another.

5 CATALOGUE PROPERTIES

The final BGS catalogue in the DECam region in the SGC covers the declination range $-17^\circ \lesssim$ Dec. $\lesssim 32^\circ$, and in the NGC the range $-10^\circ \lesssim$ Dec. $\lesssim 32^\circ$. The BGS has a total of 13 378 062 galaxies of which 7944 975 are in BGS BRIGHT and 5433 087 are in BGS FAINT. The total area covered by the BGS in the DECaLS subregion defined by the footprint of the tiles in Fig. 1 and after accounting for the spatial cuts described in Section 3 is 9401 deg^2. In Table 2, we list the target density of the BGS catalogue for each of the best-fitting photometric models used in TRACTOR.

In Fig. 8, we show the BGS BRIGHT and BGS FAINT sky map densities computed with the HEALPIX scheme using

$$\begin{aligned} \eta_i &= N_i^{\rm BGS}/A_{\rm eff}, \\ A_{\rm eff} &= N_i^{\rm R}/\eta^{\rm R}, \end{aligned} \tag{7}$$

where for each pixel $N_{\rm i}^{\rm BGS}$ is the number of BGS targets, $A_{\rm eff}$ is the effective area computed from the number of randoms, $N_i^{\rm R}$, and the total surface density of the randoms, without any masking, is $\eta^{\rm R}$

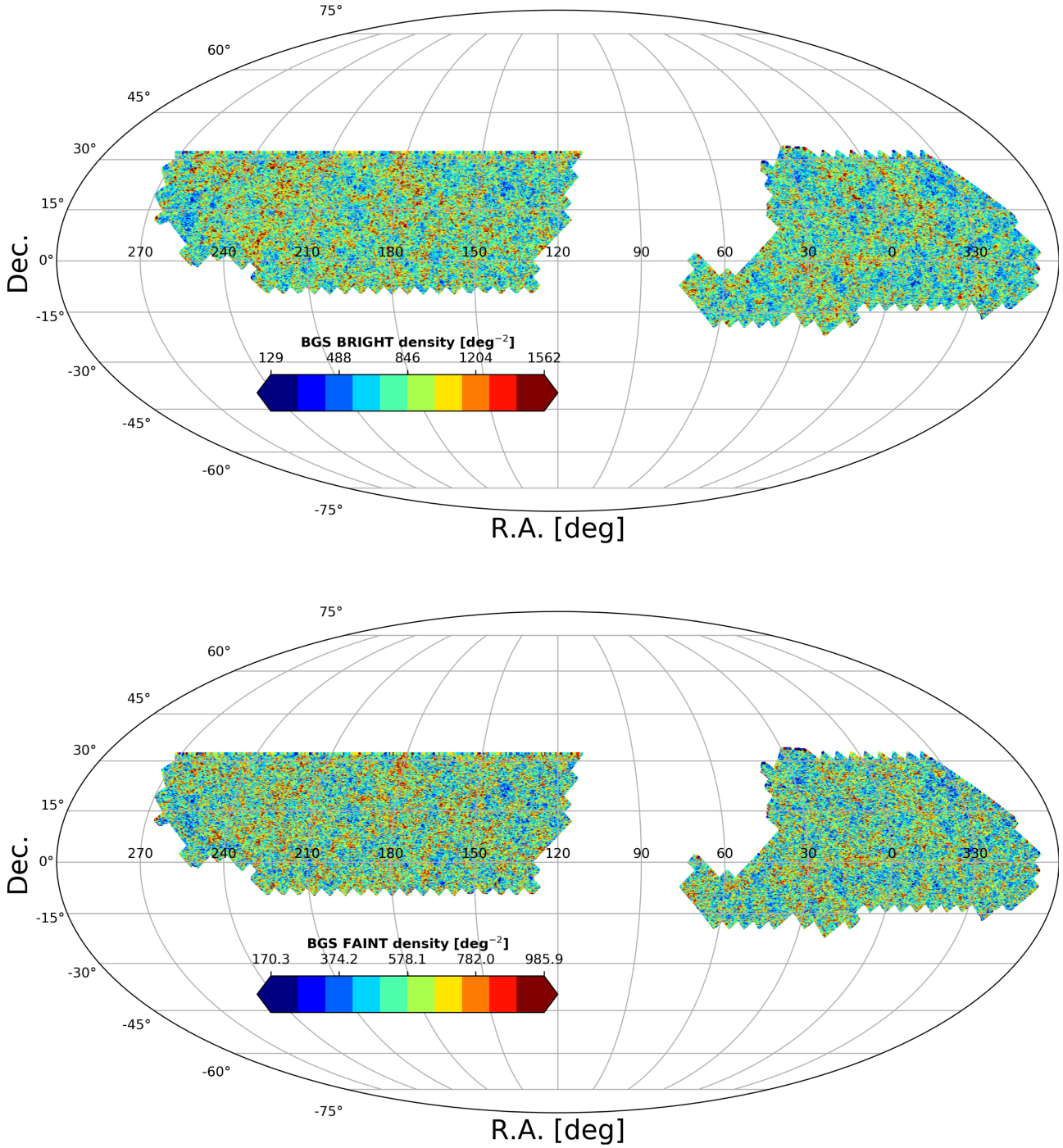

Figure 8. The distribution on the sky of the BGS BRIGHT (upper map) and BGS FAINT (lower map) target density in objects deg^{-2}, computed on a HEALPIX grid with a resolution of $N_{\rm side} = 256$. The mean densities are 846 and 579 objects deg^{-2} for the bright and faint BGS, respectively.

$= 15\,000$ objects deg^{-2}. We use a grid of $N_{\rm side} = 256$ giving a pixel area of $A_{\rm pix} = 0.052$ deg^2. The appearance of the density fluctuations is very similar in the two disjoint regions and shows no variation with galactic latitude. We look more closely at systematic variations in the target density in Section 5.2.

5.1 Cross-comparison with GAMA

The main target sample in GAMA (Baldry et al. 2017) is a complete sample of galaxies with SDSS Petrosian r-band magnitude brighter than $r = 19.8$. The Petrosian magnitude is measured within a circular aperture of twice the Petrosian radius, where the radius is computed using the r-band surface brightness profile (Adelman-McCarthy et al. 2008). The GAMA photometric selection is very similar to that of DESI BGS and so we expect a very similar redshift distribution as GAMA that has median of $z = 0.2$ and a 90 percentile value of $z = 0.5$.

Star–galaxy separation in GAMA was conservative in that it aimed for very high completeness at the expense of some stellar contamination. These properties combined with its very high spectroscopic completeness (high-quality redshift have been obtained for more than

98.85 per cent of the GAMA targets) make it a nearly ideal 'truth table' from which to assess the completeness of the BGS target selection and measure the expected redshift distribution of the BGS BRIGHT sample. Below we make use of GAMA to examine various aspects of our BGS catalogue. In Section 5.1.1, we compare the *r*-band photometry of the matched objects and determine the redshift distribution of the BGS galaxies that match with galaxies in the GAMA survey. Section 5.1.2 explores an issue related to TRACTOR only providing PSF photometry for some of the BGS galaxies. In Section 5.1.3, we assess incompleteness in BGS relative to GAMA and quantify how much is caused by each of the various geometric and photometric selections.

5.1.1 Magnitude definition and redshift distribution

We match the GAMA Main Survey DR4 galaxy catalogue (Driver et al. 2011; Liske et al. 2015; Baldry et al. 2017), which is defined by a Petrosian magnitude (R_PETRO) limit of $r = 19.8$, to the BGS target catalogue. We use a maximum linking length of 1 arcsec to match them. The mean separation of the matches we find is 0.093 arcsec with a 1σ dispersion of 0.091 arcsec. We focus on three of the five GAMA fields: G09, G12, and G15. We omit G02 as this GAMA field is only partially within the DECaLS footprint, and G23 is far to the south. The redshift completeness of the main GAMA survey is extremely high in the sense that 98.85 per cent of the objects in the catalogue yield redshifts with a quality flag NQ ≥ 3.

The GAMA spectroscopic redshifts can be used to reliably reject stars with a cut at $z = 0.002$. In what follows we restrict our GAMA catalogue to the spectroscopically confirmed galaxies ($\sim$98 per cent of the full catalogue). The area of each of the GAMA fields considered is 59.98 deg^2 that means that our matched sample has a total area of $\sim$180 deg^2. The overall density of sources that are cross-matched between BGS and GAMA galaxies is $\sim$970 objects deg^{-2} with a mean redshift of $z = 0.224$.

For this matched catalogue, Fig. 9 compares the DR8 *r*-band total magnitude ($r_{\rm LS}$) with the Petrosian *r*-band magnitude from GAMA ($r_{\rm GAMA}$) by plotting $r_{\rm LS} - r_{\rm GAMA}$ versus $r_{\rm GAMA}$. To see how this difference depends on galaxy morphology, we divide the LS galaxies into the five photometric classes assigned by TRACTOR. In each panel we show the fraction of matched galaxies in each TRACTOR model fit class; DEV and EXP classes together make up 80 per cent of the sample and the PSF class just 2.5 per cent. We mark on the plot the $r_{\rm LS} < 20$ limit of BGS, but note this has not been applied when defining the LS sample that was matched to GAMA.

Differences in the effective passbands of the *r*-band filters of the LS and SDSS result in offsets in $r_{\rm LS} - r_{\rm GAMA}$ of around -0.05 and -0.1 for blue and red galaxies, respectively (Dey et al. 2019). One also has to consider the difference in magnitude definitions that contributes the more to this magnitude offset. To the extent that the best-fitting profiles accurately describe the actual light profiles of the objects, LS provides total magnitudes. In contrast, the SDSS Petrosian magnitudes used by GAMA quantify only the flux within twice the Petrosian radius (Blanton et al. 2001). The fraction of the flux within this aperture depends on the light profile. For EXP profile it captures 99.4 per cent, but for the DEV profile, which is more sharply peaked but with broader wings, only 82 per cent is captured. It is these differences in definition that largely drive the differences in median offsets we see in the DEV, EXP, REX, and COMP classes. In all these cases the LS magnitude is brighter (more negative) than the GAMA magnitude with median offsets being -0.085 mag for EXP and -0.188 mag for DEV. In contrast, for the PSF case the median $r_{\rm LS} - r_{\rm GAMA}$ is positive, which means that the LS PSF model magnitude captures less flux than the GAMA Petrosian magnitude. For true point sources we would expect these two magnitudes to be almost equal. The positive difference appears to happen because TRACTOR force fits PSF models to sources that are actually extended (deemed extended by our *Gaia*-based star–galaxy separation) and consequently underestimates their fluxes. The reason this happens is discussed in Section 5.1.2.

If we take account of the scatter between the BGS and GAMA magnitudes, we can use GAMA to assess the level of contamination in the BGS catalogue. If we treat GAMA as being a 100 per cent complete galaxy catalogue, then any objects in BGS that are not in GAMA would be contamination in the form of stars or image artefacts. This is not true at $r = 20$ as here some BGS objects will not be in GAMA simply because of the $r_{\rm petro} < 19.8$ mag limit in GAMA. This can be seen in Fig. 9 from the location of the $r_{\rm LS} = 20$ dashed line relative to where the GAMA data truncate at $r_{\rm GAMA} = 19.8$. To avoid this problem if we apply a brighter magnitude limit $r < r_{\rm lim}$ to BGS, then for a broad range of $18.5 \lesssim r_{\rm lim} \lesssim 19.3$ we find that $\sim$3 per cent of BGS objects are not matched with GAMA galaxies. This sets an upper limit (in this magnitude range) of 3 per cent contamination in BGS as GAMA itself may not be 100 per cent complete.

Fig. 10 shows the distribution of redshifts for BGS objects that have been cross-matched with GAMA galaxies. The overall distribution is shown along with those for the BGS FAINT and BGS BRIGHT. We expect this distribution to be representative of the BGS BRIGHT sample as we can see from Fig. 9 that incompleteness caused by the GAMA magnitude limit to be very small. However the redshift distribution plotted for BGS FAINT is more strongly affected by the GAMA magnitude limit and its true redshift distribution is expected to be more extended.

5.1.2 Galaxies with TRACTOR-type PSF

To avoid stars being classified as extended sources TRACTOR uses a catalogue of stars from *Gaia* to pre-select a set of objects on which it will only allow PSF fits. The *Gaia* objects for which it does are based on the following cut on the *Gaia* astrometric excess noise (AEN) parameter,

$$\begin{aligned} \mathrm{AEN} &< 10^{0.5}, \quad G \leq 19, \\ \mathrm{AEN} &< 10^{0.5+0.2(G-19)}, \quad G \geq 19, \end{aligned} \tag{8}$$

where G is the *Gaia* photometric G-band. The AEN can be used as measure of whether a source is extended as for extended sources the astrometric measurements are noisier than one would expect for a point source.

In contrast, in the BGS we use the difference between the *Gaia* G-band magnitude and the TRACTOR raw *r*-band magnitude, rr (not corrected for extinction), as a measure of how extended the object is (see Section 4.1). In Fig. 11, we have plotted log (AEN) versus G separately for objects classified as stars and galaxies by our $G - \mathrm{rr}$ classifier. The threshold adopted by TRACTOR can be seen to separate the bulk galaxies from the stars. For 96 objects deg^{-2} the two criteria agree the object is a galaxy, but the distributions are extended and the agreement is not perfect. There are 36 objects deg^{-2} that the AEN criterion classifies as galaxies that $G - \mathrm{rr}$ classifies as stars. More problematic are the 5 objects deg^{-2} that the AEN criterion classifies as stars that $G - \mathrm{rr}$ classifies as galaxies. This is an issue as it means some objects that are classified as galaxies in the BGS are treated by TRACTOR as stars and only have a PSF light profile

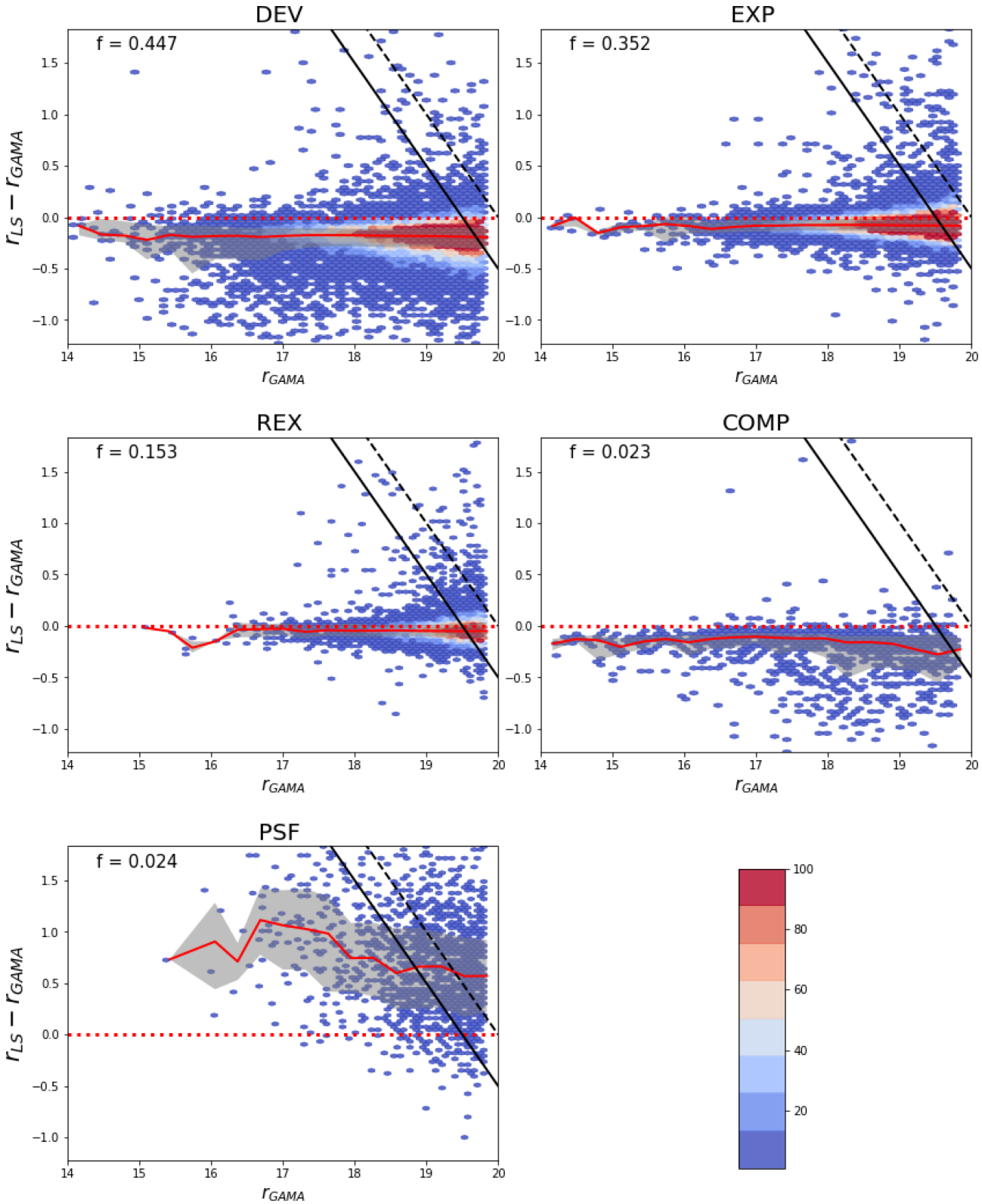

Figure 9. The *r*-band total magnitude in the LS ($r_{\rm LS}$) versus the SDSS *r*-band Petrosian magnitude in GAMA ($r_{\rm GAMA}$) for LS DR8 objects cross-matched with GAMA. Each plot corresponds to one of the five photometric model fits assigned by TRACTOR. The red solid line shows the median value of $r_{\rm LS} - r_{\rm GAMA}$ as function $r_{\rm LS}$; the grey shading shows the 20–80 percentile range; the dashed black line shows the limiting magnitude of $r_{\rm LS} = 20$ for BGS; and the solid black line shows limiting magnitude of $r_{\rm LS} = 19.5$ for BGS. The colour bar shows the number counts of objects in a hexagonal cell covering the range from. The fraction of LS DR8 objects plotted out of the total number matched with GAMA are shown in the top left-hand corner of each panel.

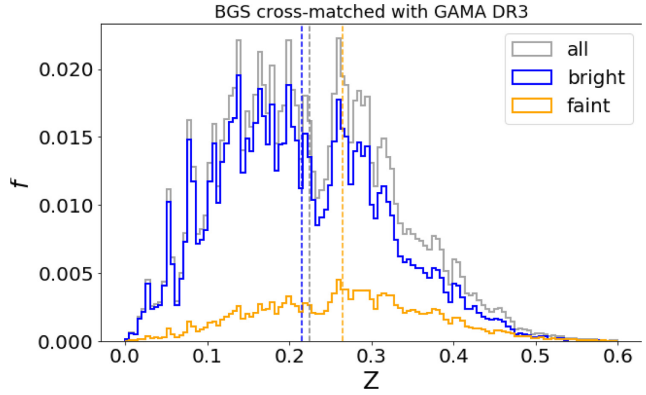

Figure 10. The redshift distribution of BGS objects cross-matched with GAMA DR4 broken into bright ($r < 19.5$, blue) and faint ($19.5 < r < 20$, orange) galaxies according to the BGS r-band. The grey histogram shows the overall redshift distribution of BGS galaxies cross-matched with GAMA. The mean redshift values for each distribution are: 0.215 for the bright sample (dashed blue line); 0.265 for the faint sample (dashed orange); and 0.224 for all galaxies (dashed grey).

fitted. Overall in the BGS there are 5 objects deg^{-2} with PSF type within the DECaLS footprint (see Table 2). These objects have fibre magnitudes that are consistent with the locus of stars in Fig. 6 that makes us question if they really are galaxies. We investigate this below by making use of GAMA to determine whether or not they are galaxies.

First, we restrict our attention to the 180 deg^2 of our matched GAMA catalogue. The BGS PSF-type galaxies (main sample) have a density of 4.10 objects deg^{-2}, somewhat less than the 5 objects deg^{-2} that is the average over the full DECaLS area. This reduces further to 1.76 objects deg^{-2} after cross-matching with GAMA. We further subdivide these two cases (BGS PSF type and BGS PSF type cross-matched with GAMA) into three disjoint subsamples: (i) those that are not in *Gaia*; (ii) those that are in *Gaia* and that are classified using the AEN value as stars; and (iii) those that are in *Gaia* and that are classified using the AEN value as galaxies.

The subsample sizes are reported in Table 3, where we give the surface density of objects before and after the cross-match with GAMA ($\eta_{\rm BM}$ and $\eta_{\rm AM}$) along with the percentage of the total number of objects represented by each subsample. This shows that ∼96 per cent of the BGS PSF type cross-matched with GAMA are *Gaia* AEN stars, which represents the ∼55 per cent in the non-matched sample. For the remained 45 per cent in the non-matched sample, GAMA is not reliable to assess this as only 3.6 per cent of those are cross-matched with GAMA. Fig. 12 shows the GAMA redshift distribution for the BGS PSF type cross-matched with GAMA broken into the three classes shown in Table 3. These objects show a redshift distribution very similar to that of the full BGS sample. The reason for this misclassification lies in the fact that for objects classified by the *Gaia* AEN criterion as stars TRACTOR only fits PSF models. For the galaxies that this *Gaia* AEN criterion falsely classifies as stars TRACTOR underestimates the total flux of the galaxy resulting in the offset with the GAMA photometry we saw in the PSF panel of Fig. 9 and putting these galaxies close to the stellar locus in Fig. 6.

5.1.3 Incompleteness of BGS relative to GAMA

To the depth of GAMA we can assess the completeness of the BGS catalogue by cross-matching the full depth LS DR8 catalogue with GAMA DR4. This cross-match yields a catalogue of 1011 objects deg^{-2} that represents of 99.6 per cent of the GAMA catalogue. Visual inspection reveals some of the remaining 0.4 per cent are deblending issues where GAMA fragments a galaxy into two objects while TRACTOR keeps it as a single object. Of the matched objects 970 objects deg^{-2} are in BGS, while the other 41 objects deg^{-2} are excluded from the BGS catalogue by one or other of our selection cuts.

Because of the scatter between SDSS r-band Petrosian magnitude used by GAMA and the TRACTOR model magnitude used by BGS (see Fig. 9), the BGS $r_{\rm LS} = 20$ mag limit excludes 20 faint GAMA galaxies per deg^2. This leaves 20.8 objects deg^{-2} in GAMA that are missing from the BGS. Whether this represents potential problematic incompleteness in BGS or just a difference in sample definition depends on which selection cuts remove the objects. We quantify and discuss this below.

The diagonal elements in Fig. 13 indicate the number density of spectroscopically confirmed GAMA galaxies missing from the BGS catalogue as result of each of the following spatial and photometric cuts: the bright star (BS) mask; the large galaxy (LG) mask; the number of observations (NOBS); star–galaxy (SG) classification; fibre magnitude cut (FMC); colour cuts (CCs); the FRACMASKED quality cuts (QCsFM); the FRACIN quality cuts (QCsFI); and the FRACFLUX quality cuts (QCsFF). The off-diagonal entries in Fig. 13 show the surface density of GAMA galaxies that are removed by both of the two cuts indicated by the labels on the x and y axes.

The objects removed by the spatial BS and NOBS cuts are benign in that they do not affect BGS clustering measurements. These spatial masks are uncorrelated with BGS galaxy positions and so can be fully accounted for in clustering analyses by applying the same masks to the random catalogue. The values given in Fig. 13 show that these two masks have no overlap and together remove 9.36 objects deg^{-2}.

Applying these two spatial cuts leaves us with 11.43 galaxies deg^{-2} that are in GAMA but are missed by BGS. The cuts that remove these objects are almost completely independent. 5.36 objects deg^{-2} are removed by our SG classification. These objects are close to the cut imposed for the *Gaia* star–galaxy separation ($G -$ rr $= 0.6$), but fall on the stellar side. We find that 98 per cent of these missed GAMA galaxies are classified as stars according to the *Gaia* AEN condition, which means that their photometry has been compromised as TRACTOR only fitted PSF models. If these are extended objects, then their flux as reported by TRACTOR is a fraction of what it should be and hence their rr magnitude is shifted to fainter values. This results in BGS galaxies shifting to lower values of $G -$ rr, moving them out of the galaxy locus and into the stellar one. If this were fixed we would expect the residual incompleteness to be 6.07 galaxies deg^{-2}, equivalent to $6.07/970 = 0.62$ per cent. The proportions of this produced by the LG QCsFM, QCsFI, and QCsFF cuts are 23.5, 41.2, 13.8, and 21.4 per cent, respectively, with a negligible fraction removed by the CCs and FMC.

In Fig. 14, we show the redshift distribution of the GAMA galaxies that are not present in the BGS. The solid coloured lines show the distribution for GAMA galaxies rejected by different BGS cuts, as labelled in the figure. We also plot the overall redshift distribution of BGS galaxies for comparison. GAMA galaxies removed by the bright star masking and by the restrictions on the number of observations have a similar redshift distribution to the overall BGS. GAMA galaxies that are removed by the large galaxy mask have a distribution that is shifted to lower redshifts than the overall BGS distribution. GAMA galaxies can be found within the geometric BGS mask as GAMA does not use masking to deal with large galaxies, and so GAMA galaxies can be found in the regions that the

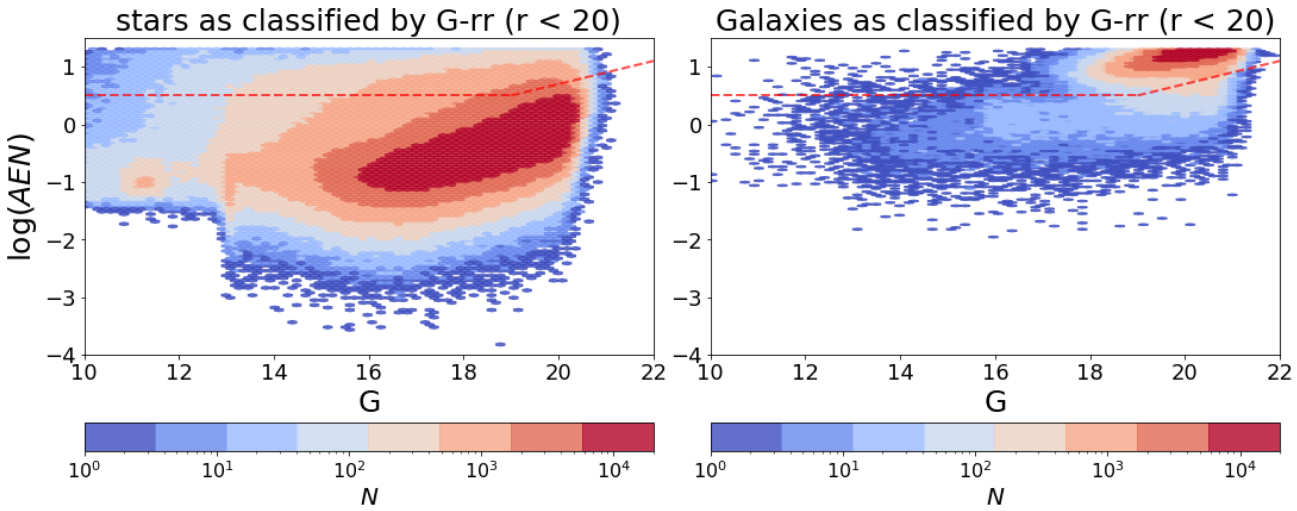

Figure 11. The *Gaia* astrometric excess noise (AEN) parameter versus *G*-band magnitude. The left-hand panel shows *Gaia* objects classified as stars by BGS and the right-hand panel shows those classified as galaxies. Both plots only show *Gaia* objects with magnitudes limit of $r < 20$. The red-dashed line represents the threshold limit for the AEN classification used in TRACTOR, therefore everything below the line is a star and everything above is a galaxy according to the AEN classification. The colour bar shows the number counts of objects in a hexagonal cell covering the range from 1 to 20 000.

Table 3. The surface density of PSF-type objects in the BGS in the G09, G12, and G15 GAMA fields combined before (η_{BM}) and after (η_{AM}) cross-matching with GAMA (top half of table). The bottom half of the table shows the surface density and percentage of objects in disjoint subsamples of the PSF-type BGS sample, as listed in the first column: objects that are not in *Gaia*, objects that the AEN scheme classifies as stars, and those that the AEN scheme classifies as galaxies.

Sample	η_{BM} (deg^{-2})		η_{AM} (deg^{-2})	
PSF-type BGS	4.10		1.76	
Subsample	η_{BM} (deg^{-2})	Per cent$_{BM}$	η_{AM} (deg^{-2})	Per cent$_{AM}$
Not in *Gaia*	1.72	42.0	0.04	2.3
Gaia AEN star	2.26	55.2	1.69	96.4
Gaia AEN galaxy	0.11	2.8	0.02	1.3

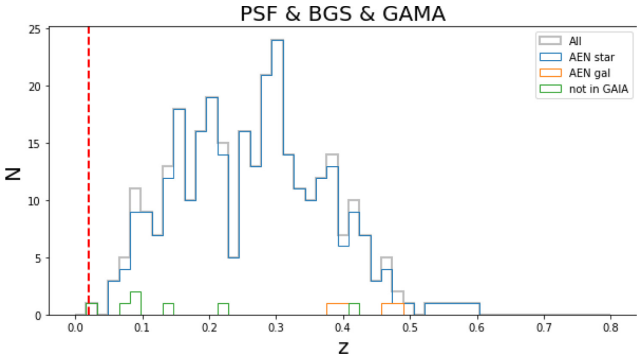

Figure 12. Redshift distribution of PSF-type BGS galaxies cross-matched with galaxies from three GAMA fields (G09, G12, and G15). Redshifts are taken from GAMA DR4. The four distributions correspond to the matched sample (grey) and the disjoint subsamples comprising galaxies not in *Gaia* (green), and stars (blue) and galaxies (red), as defined by the AEN classification. The red dashed line marks the redshift $z = 0.002$; objects with redshifts smaller than this are stars.

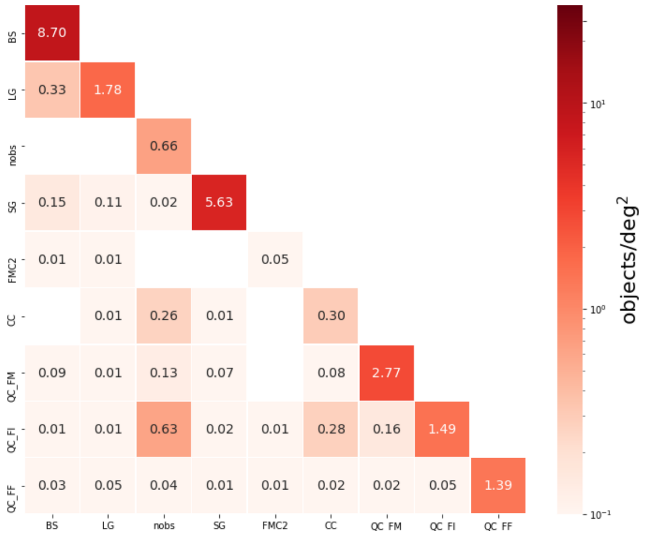

Figure 13. Heat map showing the target density of GAMA galaxies ($z > 0.002$) that are missed in the BGS. The diagonal shows the number of objects per deg^2 removed by each of the individual spatial and photometric cuts applied in the BGS, while the off-diagonal entries show the densities of objects removed by both cuts labelled on the *x* and *y* axes.

BGS rejects around large galaxies. However, GAMA does perform masking around bright stars but this is less aggressive than the LS DR8 bright star masking. This can be seen from the areas rejected: the bright star masking in GAMA removes ∼1 object deg^{-2} (Baldry et al. 2010), whereas LS DR8 removes ∼5 objects deg^{-2}.

5.2 Potential systematics

Here we look at potential systematic effects that could influence the homogeneity of the BGS catalogue and show how to mitigate these. As in any survey, the density of BGS targets is affected by observational effects that arise for a number of reasons. These include astrophysical foregrounds such as Galactic extinction, variations in

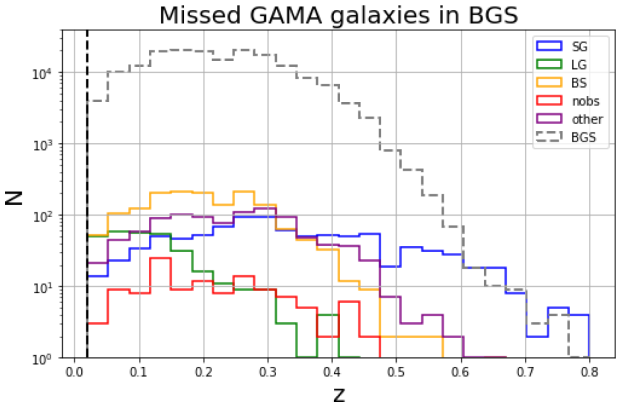

Figure 14. Redshift distribution of the GAMA galaxies that are not included in the BGS, with objects rejected by different cuts indicated by different line colours as labelled: blue shows GAMA objects missed due the star–galaxy separation applied (SG), green due to large galaxy masking (LG), yellow – bright star masking (BS), red – number of observations (NOBS), and purple due to the remaining cuts (CCs, FMC, and all the QCs). The dashed grey line shows the redshift distribution of BGS galaxies cross-matched with GAMA. The vertical black dashed line marks the redshift boundary between stars ($z < 0.002$) and galaxies.

the density of stars in the Milky Way, and variations in depth for the different imaging surveys and uncertainties in the data calibration.

To study the impact of these systematics on the observed galaxy density, we use a HEALPIX map that divides the whole sky into $12N_{\rm side}^2$ equal area pixels, adopting $N_{\rm side} = 256$. Each pixel contains the median value of the systematics values within the pixel and the BGS target density. The corresponding BGS target density in each pixel, η_i, is defined in equation (7).

We study the effect of eight systematics on the BGS target density.

(i) Stellar density: we use stars from the *Gaia* DR2 catalogue with $12 < G < 17$ to construct the stellar density in each HEALPIX pixel.
(ii) Galactic extinction: the extinction values were computed using the SFD98 dust maps as reviewed in Section 2.1.
(iii) PSF size (seeing) in the *grz* bands: the PSF size measures the full width at half-maximum (FWHM) of the PSF that determines how much the transmission of light through turbulence in the Eath's atmosphere blurs the observed images. The seeing varies across the multiple observations.
(iv) Photometric depth in the *grz* bands: the depth of the photometry, as characterized by the 5σ AB magnitude detection limit for a 0.45 arcsec round exponential galaxy profile, varies across the survey due to changes in the observing conditions.

To determine if the BGS target density has a systematic dependence on any of these quantities, we bin the HEALPIX pixels according to the value of the quantity and for each bin determine the mean target density, η_i, and the error on the mean, $\sigma_i/\sqrt{N_i}$. In Fig. 15, we show how the mean BGS target density, η, varies with respect to each of the quantities listed above. Each panel shows the mean and error on the mean for three samples, BGS BRIGHT, BGS FAINT, and the combined BGS sample (labelled bgs_any). The histogram below the curves in each panel shows (on an arbitrary scale) the number of HEALPIX pixels contributing to each estimate. In general, the systematic variation in the BGS target density is less than 5 per cent, with the one exception being a ∼7 per cent decrease in the target density in regions of high stellar density.

Stars could impact the BGS target density in at least five ways. (i) Stellar contamination of the BGS selection could lead to increased target density in regions of the sky with high stellar density. (ii) While the impact of very bright stars is dealt with by masking (see Section 3.1.1), the haloes and diffraction spikes around slightly fainter stars could still affect the photometry of neighbouring galaxies. (iii) High stellar density could lead to an overestimate of the local sky brightness that, when subtracted, would lead to fainter galaxy fluxes and hence a lower BGS target density. (iv) Star/galaxy superposition. (v) Binary stars that TRACTORs resolution is not capable of resolving.

Stellar contamination would lead to an increase in target density with increasing stellar density, whereas we see a decrease that sets in above a stellar density of 10^3 deg^{-2}. Hence, stellar contamination cannot be the dominant systematic influence on the target density.

Galaxy photometry directly compromised by nearby stars that were not subject to masking also seems unlikely to be the cause for the variation in target density. We test this by implementing the medium bright stars mask with a very little impact on target density and clustering. A further masking with two and three times the masking radius of equation (3) was also tested with no improvement on target density at high stellar densities.

The effect of high stellar density on the estimation of the sky levels deserves further investigation, but is deferred to another study. There is some variation of the target density with galactic extinction that could indicate systematic errors in the estimation of the amount of dust extinction. However, as there are spatial correlations between stellar density and dust extinction, these trends could be driven by the variation in stellar density and can be mitigated with several techniques such as linear and non-linear regressions and machine learning techniques such as artificial neural networks (Rezaie et al. 2020).

Because of variations in observing conditions, the PSF size varies across the survey. The explicit modelling of the PSF of each image by TRACTOR should make the photometry robust to this variation. Also, our use of *Gaia* to perform star–galaxy separation should also make this classification independent to variations in the seeing. This appears to be borne out by the results shown in Fig. 15 that exhibit only very weak trends with PSF.

In the BGS, while the primary selection is in the *r*-band, TRACTOR simultaneously fits objects in all three bands and so the model parameters are affected by data in all three bands. However, any dependence on the depth of the photometry appears very weak in all three bands. This to be expected as the photometric depth is typically 3–4 mag deeper than the $r = 20$ selection limit of the BGS.

5.2.1 Mitigation of systematics using linear weights based on stellar density

One way to mitigate the effect of the systematics in our catalogue is to apply a weight that corrects the target density. If we treat the systematic dependence of the observed target density on a particular quantity, S, as a simple regression problem, we can define the observed target density, $\eta_i^{\rm o}$, averaged over HEALPIX pixels with a particular value of $S = S_i$, as

$$\eta_i^{\rm o} = \eta_i \, W_i(S_i). \tag{9}$$

Here, η_i is the true target density and $W_i(S_i)$ is the weight for a given systematic attribute, S. As shown in Fig. 15, the most important target density variation is driven by stellar density. Here, we assume that the weight is a simple linear function, $W_i(S_i) = mS_i + c$, where S_i is the stellar density, as we would expect any contamination (or

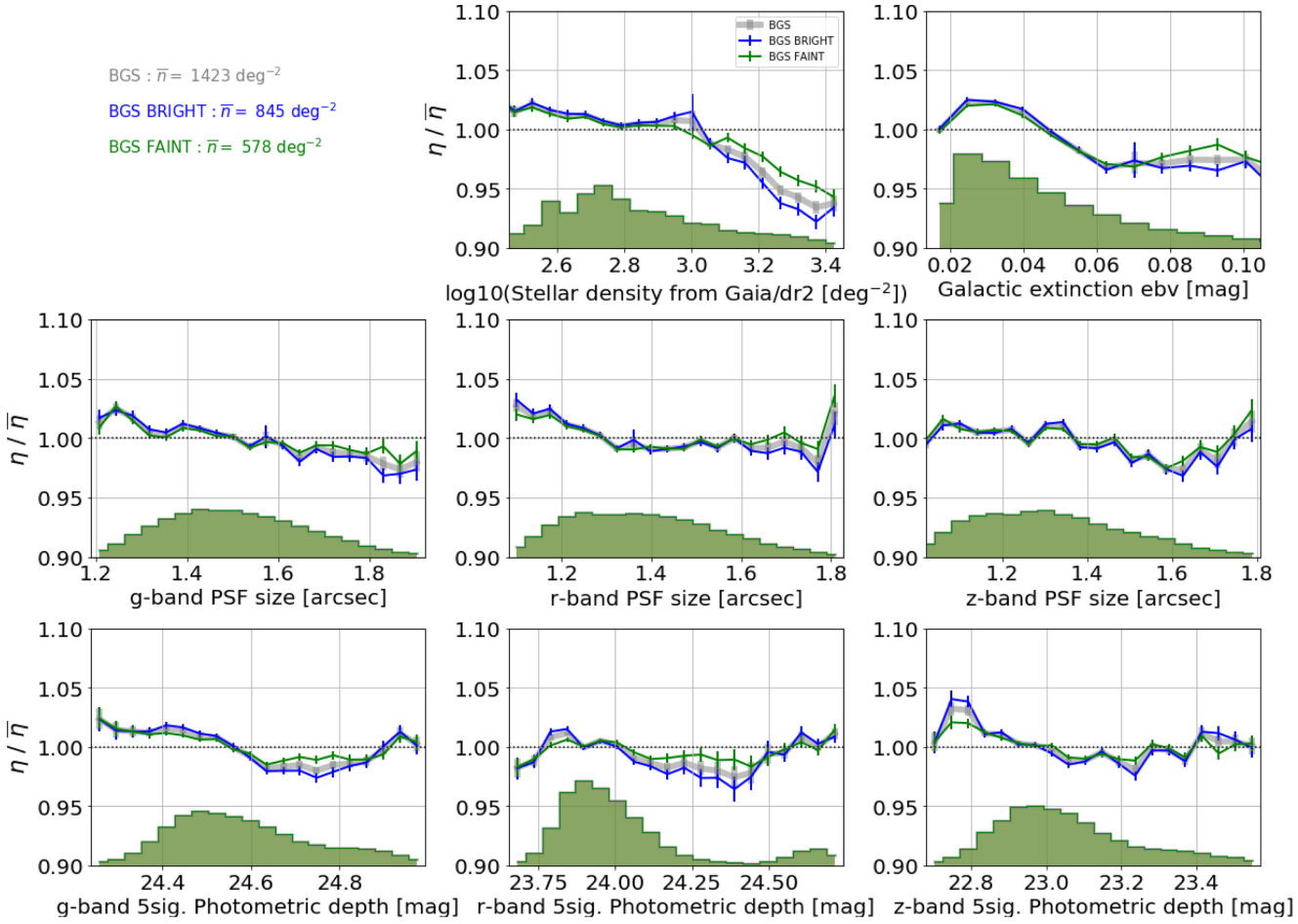

Figure 15. The systematic variation of the BGS BRIGHT (blue) and BGS FAINT (green), and combined (bgs_any, grey) target densities with respect to different properties: the logarithm of the stellar density from *Gaia* DR2; Galactic extinction; PSF size in the three bands (*grz*); and the photometric depth in each of the three bands (*grz*). The target densities and these eight quantities were computed in pixels on the sky using a HEALPIX grid with resolution of $N_{\rm side} = 256$. Histograms show the distribution for each of the *x*-axis properties. The error bars show the errors on the mean. Each target density, η, is expressed in units of its mean across the whole survey $\bar{\eta}$ as given in the legend.

anticontamination) to be proportional to the stellar density and not to the $\log_{10}$(stellar density). The best-fitting coefficients we find when applying this model to the combined BGS BRIGHT and BGS FAINT sample are $c = 1.03$ and $m = -3.96 \times 10^{-5}$. By construction, this weighting removes the general trend with stellar density for the combined sample and most of the trend with stellar density for the individual BGS BRIGHT and BGS FAINT samples. At the same time this weighting also reduces the weak systematic trend of target density with galactic extinction.

5.3 Angular correlation function

We measure the angular correlation function, $w(\theta)$, in five apparent magnitude bins from $r_{\rm AB} = 15$ to 20 for the BGS targets in DECaLS SGC and NGC. Angular correlations (see Section 5.4) were computed using the publicly available code CUTE (Alonso 2012). We compare these with measurements from the mock BGS light-cone catalogue (Smith et al. 2017). This mock catalogue was built by populating the MXXL *N*-body simulation with galaxies based on a halo occupation distribution (HOD) model. By construction, the HOD parameters of this mock reproduce both the luminosity function and two-point clustering measured in the SDSS at low redshift and the GAMA survey at higher redshift.

Fig. 16 shows the comparison of angular clustering measured for the BGS targets with error bars corresponding to the standard deviation of 100 jackknife realizations, the MXXL mock, and the SDSS observations by Wang, Brunner & Dolence (2013). The angular clustering measurements are consistent between the DECaLS-North and DECaLS-South regions, which demonstrate the homogeneity between these two parts of DECaLS. The angular clustering of the BGS targets agrees very well with that displayed in the MXXL light-cone. The HOD parameters of the MXXL mock have been tuned to attempt to match the clustering measured from SDSS MGS, however on large scales HOD models can only alter the amplitude and not the shape of the correlation. Moreover the shape of the large-scale correlation function of MXXL is very similar to that of all Lambda cold dark matter (LCDM) models that are consistent with cosmic microwave background (CMB) observations. Hence it is interesting that for the two faintest bins BGS is more consistent with MXXL (and hence with LCDM) than is SDSS MGS – possibly indicating reduced systematic errors.

We also look at the angular clustering of the BGS targets after applying the weights that depend on stellar density, as described in the previous section. Overall, applying stellar density weights has a small impact at angular scales larger than 3°–4°. Both the clustering

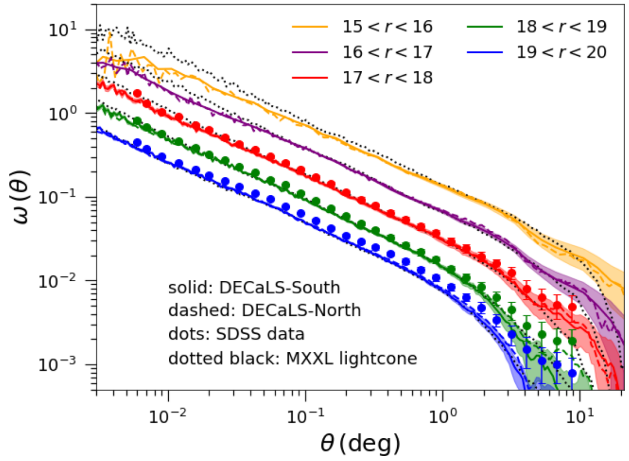

Figure 16. The angular correlation function, $w(\theta)$, measured for the BGS targets in bins of apparent magnitude; different colours indicate different magnitude bins as labelled. The shaded area shows the standard deviation obtained from 100 jackknife regions. The solid curves show the results for DECaLS-South, the dashed curves show DECaLS-North, and the dotted curves show the angular clustering in the MXXL light-cone catalogue. The symbols with error bars show measurements from the SDSS by Wang et al. (2013).

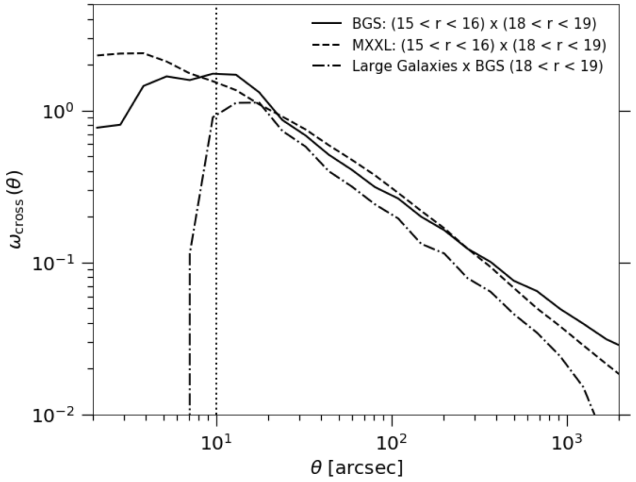

Figure 17. The angular cross-correlation function measured between faint BGS targets in $18 < r < 19$ and large galaxies from the SGA-2020 (dash–dotted) and between the same faint BGS targets and brighter BGS targets in $15 < r < 16$ (solid), the magnitude range in which most of the large galaxies reside. We also compare with the angular cross-correlation between these two bins in apparent magnitude measured in the MXXL light-cone (dashed). The vertical dotted line shows the mean LG mask radius, which is about 10 arcsec.

with and without the weights are consistent with each other, within the error bar.

A further test of the fidelity of our BGS catalogue is to check for any spatial correlation of the distribution of BGS targets with stars in the Milky Way. Here we focus our attention on the fainter stars, $12 < G < 17$, which, ideally, should be removed from the BGS targets by our star–galaxy separation scheme. We find a significant anticorrelation on very small scales but no correlation on scales larger than 100 arcsec.

5.4 Angular cross-correlation with large galaxies

In order to determine whether we are missing faint BGS targets around large galaxies due to the LG mask defined in Section 2.2.4, we measure the angular cross-correlation function between the SGA-2020 and faint BGS targets in $18 < r < 19$ (dash–dotted) as shown in Fig. 17. We also measure the angular cross-correlation function between these faint BGS targets and brighter BGS targets in the magnitude range $15 < r < 16$ (solid) where we assume that most of the large galaxies lie, and we do the same using the MXXL light-cone (dashed). The vertical dotted line shows the mean mask radius around large galaxies, which is about 10 arcsec.

The agreement between the results from the BGS catalogue (solid) and from the MXXL light-cone (dashed) suggests that our treatment of large galaxies is satisfactory and we are only missing BGS targets on scales below 10 arcsec, which is the median large galaxy masking radius (see Section 3.1.2). The difference in amplitude between the solid and dash–dotted curves, with a lower value when cross-correlating with the SGA-2020, suggests that the catalogue of large galaxies contains either more low-z galaxies or more brighter galaxies, or both, compared to the BGS targets in $15 < r < 16$.

6 CONCLUSIONS

Here we have presented the steps needed to define and select the BGS targets for the DESI project. Our galaxy selection uses DECaLS LS imaging data from DR8 reduced by the NSF's OIR LabCP and TRACTOR pipelines. Our BGS target selection has two main components, one that imposes spatial cuts and the other that applies photometric selections. Figs 2 and 4 show the flowcharts that set out these two selections. At each step these flowcharts report the remaining survey area and surface density of targets.

The main features of our spatial and photometric cuts are the following.

(i) The BGS spatial target selection removes area near bright stars (BS mask), large galaxies (LG mask), and globular clusters (GC mask), as well as galaxies with less than a specified minimum number of observations (NOBS mask). The BS mask is a circular aperture that scales with the magnitude of the bright star (see equation 3). The exclusion of areas around bright stars removes ∼270 deg^2, i.e. 2.76 per cent of initial footprint. Inspection of stacked images around bright stars (i.e. those with *Gaia* $G < 13$ or Tycho-2 $V < 13$) shows that the BS masking radius used in TRACTOR is well motivated, with no sign of contamination around the bright stars in the BGS target density. There is a modest ∼6 per cent increase in BGS target density just beyond the edge of the masked region. We find that there is a negligible angular cross-correlation between stars and galaxies at scales >100 arcsec. Below 100 arcsec we have an anticorrelation possibly caused by the stars masked within the range $12 < G < 13$.

(ii) The LG and GC masks account for a smaller number of contaminants than the BS mask, removing just ∼9 deg^2 of survey area or 0.09 per cent of initial footprint.

(iii) DECaLS DR8 is complete to 99.5 per cent with at least one observation in the three bands *grz*, as described by the value of NOBS. The selection made on NOBS removes ∼39 deg^2 of imaging data.

(iv) We use *Gaia* DR2 to separate stars and galaxies as described in Section 4.1. This classification exploits the small PSF of the *Gaia* imaging compared with that typically present in ground-based observations. In our classification scheme we compare the measurement of the flux of an object by *Gaia* with that from TRACTOR through the parameter $G - \mathrm{rr}$. Objects with a TRACTOR flux that is greater than that reported by *Gaia* are considered to

be galaxies because this difference implies that they are extended sources (see Fig. 5).

(v) A small fraction (∼0.35 per cent) of BGS galaxies are of PSF type according to TRACTOR. About half of these are compact sources for which the PSF model is the best fit, but the other half have only PSF photometry as they were designated stars based on the *Gaia* AEN parameter before TRACTOR was run. For these objects TRACTOR only performs PSF fits. Matching to GAMA reveals that most (96 per cent) of these BGS PSF-type objects are confirmed to be galaxies by the GAMA spectroscopy. In addition, we find that the ∼7 GAMA galaxies deg^{-2} that are missed in BGS are mostly (∼98 per cent) of PSF type according to TRACTOR. We conclude that using the AEN classification is (i) causing ∼0.17 per cent of BGS galaxies to be falsely classified as of PSF type and (ii) compromising the photometry of another 7 objects deg^{-2} that then due to having their fluxes underestimated are falsely classified as stars by the BGS $G - \mathrm{rr}$ star–galaxy classification.

(vi) Possible systematic effects in DECaLS leave a small imprint on surface density of BGS sources. The variation in the target density of BGS sources as a function of the main possible systematic effects, such as the stellar density, galactic extinction, seeing, and imaging depth, is less than 10 per cent in the case of stellar density and under 5 per cent for the remaining systematics. We implement a weighting scheme based on a linear regression model that uses the density of stars to mitigate these effects. Applying the resulting weights, variation in the target density with stellar density is removed by construction, and is greatly reduced when plotted against the other systematic quantities.

(vii) Angular clustering measurements made from our BGS target catalogue are compared with previous measurements from SDSS and the predictions from the MXXL light-cone mock catalogue, which on large scales can be taken as a prediction of LCDM models (see Section 5.3). On small scales, the three measurements of the angular correlation function agree well, with the exception of the brightest galaxies considered. At large scales, the angular clustering we find for the BGS targets is closer to that recovered from the MXXL mock catalogue than the SDSS measurements. The agreement between the BGS and the MXXL light-cone is even better after applying the linear weights based on stellar density to the BGS.

Galleries with examples of BGS targets divided in BGS BRIGHT and BGS FAINT can be found at http://astro.dur.ac.uk/~qmxp55/bgs_ts_paper_gallery.html along with galleries showing examples of rejected objects by the different spatial and photometric cuts we apply in BGS. We included also examples of discrepancies between our star–galaxy (SG) classification using *Gaia* with TRACTORs divided into (1) TRACTORs extended objects that fail our SG classification, and the TRACTORs point sources objects that pass our SG classification, (2) are *Gaia*, and (3) are not *Gaia* sources. Finally, examples of discrepancies between TRACTORs point source classification for *Gaia* objects and our SG classification divided in two samples: (1) are galaxies by our SG classification but stars according to TRACTORs assessment of *Gaia* sources using the AEN parameter from *Gaia*, and (2) stars by our SG classification but galaxies by their AEN classification.

In a second paper, we will focus on applying this framework to select BGS targets using the additional LS, BASS, and MzLS imaging data, and set out what is needed to tune our selection to use the upcoming release of the LS DR9. Among the main changes in DR9 compared to DR8 are (i) the implementation of an iterative source detection process in TRACTOR in which the detection algorithm is rerun after sources have been fitted and subtracted, (ii) an extended PSF model to subtract the wings of bright stars, (iii) the COMPOSITE (COMP) TRACTOR model has been replaced with a SERSIC (SER) model, (iv) the criteria used to determine which *Gaia* objects are forced to be fitted by the PSF model are now more restrictive, (v) adjustments have been made to the masking procedure around bright stars and to fainter MEDIUM stars where the masking radius around bright stars has been reduced by a factor of 2. In addition, TRACTOR implements a local fit to the sky background around these objects. (vi) SGA-2020 and GC catalogues have been updated and the large galaxy photometry redone in their own custom run of TRACTOR. It is expected that (i) will marginally increase the completeness of the BGS catalogue, (iv) will reduce the incidence of galaxies being misclassified as stars, and the other changes will improve the photometry. A second paper will quantify these changes and focus predominately on changes in selection relative to this DR8 selection. Hence most of the details of the BGS selection will be in this paper only. Despite these improvements in the quality of the selection, early test releases of the upcoming DR9 data suggest that BGS targets will not vary more than 5 per cent compared to present selection with DECaLS DR8. The second paper will also include a more complete clustering analysis using mock catalogues and colour-based clustering measurements, and a more sophisticated technique for the mitigation of systematic effects. In third paper, we will cover the work we have undertaken to define and select the BGS targets for the survey validation programme. This series of papers is intended to be complementary work to the overall DESI key project paper on target selection aimed to be released in 2021.

ACKNOWLEDGEMENTS

We acknowledge helpful conversations with Anand Raichoor and Christophe Yeche. OR-M is supported by the Mexican National Council of Science and Technology (CONACYT) through grant no. 297228/440775 and funding from the European Union's Horizon 2020 Framework Programme under the Marie Sklodowska-Curie grant agreement no. 734374. SC, PN, PZ, CMB, and JL acknowledge support from the Science and Technology Facilities Council through ST/P000541/1 and ST/T000244/1. ADM was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0019022. JM gratefully acknowledges support from the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC002008 and from the National Science Foundation under grant AST-1616414.

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under contract no. DEAC02-05CH11231. This work also made extensive use of the NASA Astrophysics Data System and of the astro-ph preprint archive at arXiv.org. Authors want to thank the GAMA Collaboration for early access to GAMA DR4 data for this work. Some of the results in this paper have been derived using the HEALPY and HEALPIX package. We acknowledge the usage of the HyperLeda data base (http://leda.univ-lyon1.fr).

The Siena Galaxy Atlas was made possible by funding support from the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC002008 and from the National Science Foundation under grant AST-1616414.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding

for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, and ST/R002371/1, Durham University, and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under contract no. DE-AC02-05CH1123, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under contract no. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory, the Science and Technology Facilities Council of the United Kingdom, the Gordon and Betty Moore Foundation, the Heising-Simons Foundation, the French Alternative Energies and Atomic Energy Commission (CEA), the National Council of Science and Technology, Mexico, the Ministry of Economy of Spain, and by the DESI Member Institutions. The authors are honoured to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

DATA AVAILABILITY

The data used in this paper are publicly available. The preliminary BGS target selection described in this paper is public at https://data.desi.lbl.gov/public/ets/target/catalogs/ and detailed at https://desidatamodel.readthedocs.io. The DESI Legacy Imaging Surveys used for this work are public at https://www.legacysurvey.org/dr8/description/.

REFERENCES

Abazajian K. et al., 2003, AJ, 126, 2081
Adelman-McCarthy J. K. et al., 2008, ApJS, 175, 297
Alonso D., 2012, preprint (arXiv:1210.1833)
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Bailer-Jones C. A. L., Fouesneau M., Andrae R., 2019, MNRAS, 490, 5615
Baldry I. K. et al., 2010, MNRAS, 404, 86
Baldry I. K. et al., 2017, MNRAS, 474, 3875
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Blanton M. R. et al., 2001, AJ, 121, 2358
Carrasco J. M. et al., 2016, A&A, 595, A7
Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)
DESI Collaboration et al., 2016, preprint (arXiv:1611.00036)
Dey A. et al., 2019, AJ, 157, 168
Driver S. P. et al., 2011, MNRAS, 413, 971
Fadely R., Hogg D. W., Willman B., 2012, ApJ, 760, 15
Flaugher B. et al., 2015, AJ, 150, 150
Gaia Collaboration et al., 2016a, A&A, 595, A1
Gaia Collaboration et al., 2016b, A&A, 595, A2
Gaia Collaboration et al., 2018, A&A, 616, A1
Høg E. et al., 2000, A&A, 355, L27
Kitanidis E. et al., 2020, MNRAS, 496, 2262
Lang D., Hogg D. W., Mykytyn D., 2016, Astrophysics Source Code Library, record ascl:1604.008
Liske J. et al., 2015, MNRAS, 452, 2087
Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, A&A, 570, A13
Odewahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1992, AJ, 103, 318
Rezaie M., Seo H.-J., Ross A. J., Bunescu R. C., 2020, MNRAS, 495, 1613
Ruiz-Macias O. et al., 2020, Res. Notes AAS, 4, 187
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Secrest N. J., Dudik R. P., Dorland B. N., Zacharias N., Makarov V., Fey A., Frouard J., Finch C., 2015, ApJS, 221, 12
Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, MNRAS, 470, 4646
Strauss M. A. et al., 2002, AJ, 124, 1810
The Dark Energy Survey Collaboration, 2005, preprint (arXiv:astro-ph/0510346)
Wang Y., Brunner R. J., Dolence J. C., 2013, MNRAS, 432, 1961
Weir N., Fayyad U. M., Djorgovski S., 1995, AJ, 109, 2401
Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, J. Open Source Software, 4, 1298

APPENDIX A: GALAXY VIEW

In contrast to the approach taken in the main paper, here we present a 'galaxy' view of the BGS selection by implementing the star–galaxy separation before the other BGS cuts (with the exception of first applying the nominal BGS magnitude limit $r < 20$). The results of this exercise are shown in Fig. A1. In this view, the geometric masking does not look as aggressive as it did in Fig. 2, with the size of the rejected area and number of objects typically reduced at each step by an order of magnitude compared to what was seen in Fig. 2. The BS mask step is the stage that is the most affected by this change in order. Next is the application of the selection on NOBS that has half the effect that it did in Fig. 2. Note that the area removed by the cuts remains unchanged as this does not depend on the number of targets but is calculated using the randoms.

In addition to the changing the order in which the star–galaxy separation is applied compared to the selection criteria presented in Sections 3 and 4, we swap the FMCs and CCs with the QCs. When comparing both schemes (Figs 4 and A1), we see a high overlap between the QCs and the FMCs of $\sim$15 objects deg^{-2} that represent $\sim$2/3 the galaxies rejected by FMC in Section 4. CCs are also affected by the sequence of cuts and the rejections due to this cut is reduced by a factor of 2 in the galaxy view.

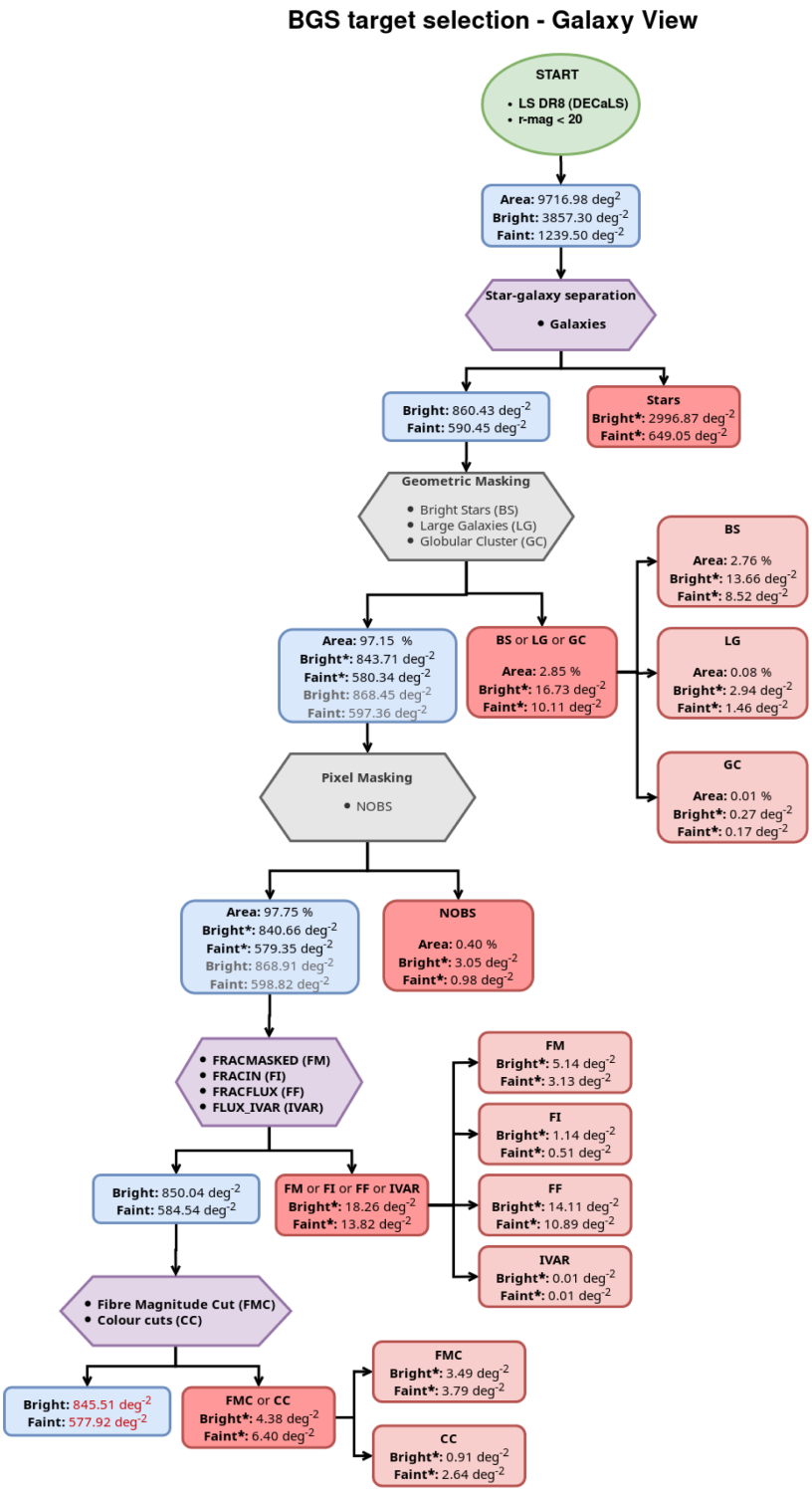

Figure A1. Flowchart showing the spatial and photometric BGS target selections applied to the Legacy Surveys DR8. The spatial selections are shown by grey boxes and are divided into two kinds: one defined by geometric cuts around bright sources, i.e. bright stars (BS), large galaxies (LG), and globular clusters (GC); and the other that is at the pixel level, such as the number of observations (NOBS). The photometric selection of BGS targets is divided into four types and is shown by purple boxes: star–galaxy separation, fibre magnitude cuts (FMCs), colour cuts (CCs), and quality cuts (QCs) that include FRACMASKED, FRACIN, FRACFLUX, and FLUX_IVAR. The blue boxes show the area (in degrees) and the number density (per deg^2) of objects retained after each selection, broken down into the numbers for the bright and faint components of the BGS. The red boxes show the equivalent information for the rejected objects. If more than one cut or selection is applied at a given stage, then the darker red boxes show the information about removed objects for the combination of cuts and the lighter red boxes show the corresponding values for each individual cut. The superscript asterisk (*) denotes target densities without correcting for the area removed by cuts up to that point, while densities without a superscript asterisk (*) do take into account the reduction in area.

This paper has been typeset from a TEX/LATEX file prepared by the author.